

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:40:24 : Search time 27 Seconds
(without alignments)

922.640 Million cell updates/sec

Title: US-09-673-400a-38

Sequence: 1 ARAPTLDMFRRLRLSADPHA.....YRGFLAEKRVQWEDEDF 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processings: Minimum Match 0%

Maximum Match 100%

List: 3 first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	397	52.0	113	11	09D1F0
2	386.5	50.7	100	11	09DC23
3	378.5	49.6	112	11	09D610
4	150	19.7	704	13	098SV9
5	137	18.0	325	4	096A68
6	137	18.0	342	4	09UPV1
7	126.5	16.6	303	4	09H7A0
8	122	16.0	371	13	093283
9	121	15.9	853	3	000833
10	105.5	13.8	1252	3	096VA3
11	101	13.2	837	3	09H7R8
12	97	12.7	231	11	09E011
13	85	11.1	639	3	09UC2
14	84	11.0	406	3	09UD9
15	83	10.9	400	6	095K1
16	79	10.4	246	16	098B8

17	78	10.2	223	2	045488
18	77.5	10.2	499	16	09A8K5
19	74.5	9.8	464	10	048967
20	73.5	9.6	589	16	09KE11
21	73	9.6	853	1	030772
22	72	9.4	183	2	093562
23	71.5	9.4	485	2	09XAF0
24	71.5	9.4	550	5	019680
25	70.5	9.2	180	16	09PAL1
26	70.5	9.2	202	4	09NV20
27	70.5	9.2	202	4	09NV20
28	70.5	9.2	443	3	006098
29	70.5	9.2	688	2	09FB28
30	70.5	9.2	797	5	09KNK3
31	70	9.2	1442	12	042066
32	69.5	9.1	476	16	09KDN8
33	69.5	9.1	1609	10	004648
34	69	9.0	763	2	09KR25
35	69	9.0	2506	12	09WJD2
36	68.5	9.0	348	2	09S213
37	68.5	9.0	471	17	09HMA9
38	68.5	9.0	558	2	09HXR0
39	68.5	9.0	1538	10	094H26
40	68	8.9	215	2	093096
41	68	8.9	250	16	0910F9
42	68	8.9	437	2	09LIX2
43	68	8.9	467	5	0967X8
44	68	8.9	940	11	035157
45	68	8.9	1113	2	09L249

ALIGNMENTS

RESULT	ID	Q9D1F0	PRELIMINARY	PRT	113 AA.
AC	Q9D1F0				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	1110012005RIK	PROTEIN (RIKEN CDNA 1110012005 GENE).			
GN	1110012005RIK				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;				
RC	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Airakawa T., Hara A., Fukunishi Y., Kono H., Aichi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,				
RA	Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,				
RA	Hayashizaki Y.,				
RT	*Functional annotation of a full-length mouse CDNA collection.*				
RL	Nature 409:685-690(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				

ID	Q9D610	PRELIMINARY:	PRT:	112 AA.
AC	Q9D610:			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	2900027G03R1K PROTEIN.			
GN	2900027G03R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Akawa T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S.,			
RA	Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kiehl P., Lewis S., Matsuo Y., Nikolaou I., Pleske G., Quackenbush J.,			
RA	Schirral I.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Futuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein N.J., But C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,			
RA	Wysshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK013603; BAB28923.1;			
DR	MED: MGI:1920115; 2900027G03R1K.			
SO	SEQUENCE 112 AA; 13578 MW; 14809ACB3D505016 CRC64;			
Query Match	49.6%; Score 378.5; DB 11; Length 112;			
Best Local Similarity	64.0%; Pred. NO. 2.1e-34;			
Matches 73; Conservative 14; Mismatches 24; Indels 3; Gaps				
QY	32 MDGRVQLKALLAGF-LRPAARRNRPPEPFQSDDRRLRPFVYQTSYMFYDNTES 90			
DB	1 MDRRIKLIMLAWPELRLRG-RMONPIPPPELFDSEMKRLPEFVYQTSYMLVYDNTED 58			
OY	91 NDALVTFELTRTPALQWVPIYRKESPLNDYDGFLEAKMRVGEWDEDF 144			
DB	59 TDKRVMLFLTRKGRALQAMARYIQTDSPMLNNTSGFLNKKERGEWDEDF 112			
RESULT 4				
Q98SV9	PRELIMINARY:	PRT:	704 AA.	
AC	Q98SV9:			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	GAG-PROTEASE.			
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;			
OC	Tetraodontidae; Takifugu.			
OX	NCBI_Taxid=31033;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TRANSPOSON-SUSHI-SAN RETROTRANSPOSON;			
RX	MEDLINE=21321259; PubMed=114298463;			

RA Butler M., Goodwin T., Simpson M., Singh M., Poulter R.;
 RT "Vertebrate LTR retrotransposons of the Tf1/sushi group."
 RL J. Mol. Evol. 52:260-274(2001).
 DR EMBL: AF316578; AAG60684.1; -;
 DR MEROPS; A02.0PW; -;
 DR InterPro: IPR001969; Asp.-protease.
 DR InterPro: IPR001995; Asp.-prot.-retrov.
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001878; ZnF_CCHC.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; ZnF_C2HC; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE: PS50175; ASP_PROT_RETROV; 1.
 DR PROSITE: PS50013; CHROMO_2; 1.
 DR PROTEASE; Zinc-finger.
 SQ SEQUENCE 704 AA; 76619 MW; 8182ELC9AB0153B3 CRC64;

Query Match 19.7%; Score 150; DB 13; Length 704;
 Best Local Similarity 27.7%; Pred. No. 5.5e-08;
 Matches 43; Conservative 24; Mismatches 52; Indels 36; Gaps 5;

OY 16 ADPHATQNSAEAGTMDGR-VOLM-----KAL 42
 DB 4 AODAVRR-TLEARGRLQHNOLLIDINASLOSINTSVTDLSLGMAGVPSPPAEAL 62
 OY 43 LAGPLPARRMRNPIPEPTFDGTDRLPEFIVQCSYWF-VDENTFSDALKVFLIT 101
 DB 63 EVYTGVAAPMEPHVPIPERYSGEAGVCASFLLQ-CSLVEFDQLPLYPDGRKIAFVN 121
 OY 102 RLTPRALQVAPIRKESPLNDYRGFLAEMKRVF 136
 DB 122 LLSGRAQWATAVLENQTPASSFSPEFTALKRVF 156

RESULT 5

O96A68 PRELIMINARY; PRT; 325 AA.

AC O96A68;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ME3 LIKE 1 (PATERNALLY EXPRESSED GENE 10 ORF1).
 GN ME3L1 OR PEG10.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Nakamura Y., Furukawa Y.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Satoh S., Furukawa Y.;
 RT "Isolation of ME3 like gene 1."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-21218929; PubMed-11318613;
 RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,
 RA Kaneko-Ishino T., Ishino F.;
 RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene
 RT Located on Human Chromosome 7q21."
 RL Genomics 73:232-237(2001).
 DR EMBL: AB049150; BAB68387.1; -;
 DR EMBL: AB049834; BAB43951.1; -;
 DE CDNA: FLJ21125 FIS, CLONE CASO6077 (HYPOTHEICAL 32.5 KDA
 SQ SEQUENCE 325 AA; 36965 MW; 118EACFA97E2A76 CRC64;

Query Match 18.0%; Score 137; DB 4; Length 325;
 Best Local Similarity 29.4%; Pred. No. 5.8e-07;

Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

OY 11 RRLSADPHATQNSAEAGTMDGRVOLMKALLAGPLPARRMRNPIPEPTFDGTD 70
 DB 45 REQVEPTPE-DEDDIELRGA-----AAAAAPPIEEEC--PEDLPKFDGNPD 92
 OY 71 LPEFIVQCSYWFVDENTFSDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLA 130
 DB 93 LAPPMAOCQIFMEKSTDFSVDRVRCFVSMATGRAARWASAKLESHILMHYPAFM 152
 OY 131 EMKRVF 136
 DB 153 EMKRVF 158

RESULT 6

O9UPV1 PRELIMINARY; PRT; 342 AA.

AC O9UPV1;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE KIAA1051 PROTEIN (FRAGMENT).
 GN KIAA1051.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-99397452; PubMed-10470851;
 RA Kikuno R., Nagase T., Ishikawa K., Hikosawa M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 6:197-205(1999).
 DR EMBL: AB028974; BA83003.1; -;
 DR InterPro: IPR001878; ZnF_CCHC.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; ZnF_C2HC; 1.
 KW Zinc-finger.
 FT NON_TER
 SQ SEQUENCE 342 AA; 38769 MW; DC14C265B92541EB CRC64;

Query Match 18.0%; Score 137; DB 4; Length 342;
 Best Local Similarity 29.4%; Pred. No. 6.2e-07;
 Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

OY 11 RRLSADPHATQNSAEAGTMDGRVOLMKALLAGPLPARRMRNPIPEPTFDGTD 70
 DB 62 REQVEPTPE-DEDDIELRGA-----AAAAAPPIEEEC--PEDLPKFDGNPD 109
 OY 71 LPEFIVQCSYWFVDENTFSDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLA 130
 DB 110 LAPPMAOCQIFMEKSTDFSVDRVRCFVSMATGRAARWASAKLESHILMHYPAFM 169
 OY 131 EMKRVF 136
 DB 170 EMKRVF 175

RESULT 7

O9H7A0 PRELIMINARY; PRT; 303 AA.

AC O9H7A0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA: FLJ21125 FIS, CLONE CASO6077 (HYPOTHEICAL 32.5 KDA
 DE PROTEIN).
 OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matnabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohnishi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-UTERUS; LEIOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK024778; BAB14998.1;
DR EMBL: BC011679; AAH1679.1;
KW Hypothetical protein.
SQ SEQUENCE 303 AA; 32478 MW; 3882DD1459808328 CRC64;

Query Match
Best Local Similarity 16.6%; Score 126.5; DB 4; Length 303;
Matches 38; Conservative 14; Mismatches 51; Indels 13; Gaps 3;

OY 28 ARGTDGGRVQLKALLAGPL--RPARRRM-----NP1PPEPDGDTDRLEPFIYQTC 79
DB 7 ASGTGCG-----KPAERGLAGHMPSSRRPRVDFCVPSDGTGDSWLLDRFLAQG 61
OY 80 SYMFDENTSNDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLAEKRV 135
DB 62 DVMSFHFHYQDNISVNCVCEILRLTGRQAQWAPYLDGDLPLPDYDELFCQDLKEY 117

RESULT 8
OY 093283 PRELIMINARY; PRT; 371 AA.
ID 093283;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPROTEIN.
GN GAG.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-SUSHI-1CH1 RETROTRANSPOSON;
RA MEDLINE-98382517; PubMed-9714821;
RA Poulter R., Butler M.;
RT "A retrotransposon family from the pufferfish (fugu) Fugu rubripes."
RL Gene 215:241-249(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-SUSHI-1CH1 RETROTRANSPOSON;
RA Poulter R.T.M., Butler M.I.;
RL Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030881; AAC33525.1;
DR InterPro: IPR001878; Znf_C2HC.1.
DR Pfam: PF00098; Zf-CCHC.1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_C2HC.1.
KW Polyprotein; Zinc-finger.
SQ SEQUENCE 371 AA; 40540 MW; 617926FC010730AC CRC64;

Query Match
Best Local Similarity 16.0%; Score 127; DB 13; Length 371;
Matches 44; Conservative 21; Mismatches 67; Indels 46; Gaps 5;

OY 2 RAPTLDMRRRLSADPHATORSAREARGT-----MDGR 35

```

```

DB 14 RPTLSPLERRVEA--HSAQLSSLOSELTKAFPTTIGSELSSQSTTSLSLSNQ 71
OY 36 VOLKALLAG-----PLPRAARRMNP1PPEPDGDTDRLEPFIYQTC 79
DB 72 SYMFDENTSNDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLAEKRV 135
OY 80 SYMFDENTSNDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLAEKRV 136
DB 131 ELLFRRQSRFVSDKAVGFTSLADKALSWAIAVDDPLRSDYSARFRREKAVF 188

RESULT 9
OY 000833 PRELIMINARY; PRT; 853 AA.
ID 000833;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPROTEIN.
GN GAG.
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-F. SP. LYCOPERSICI 42-87;
RX MEDLINE-96132549; PubMed-8544829;
RA Anaya N., Roncero M.I.;
RT "Skipsey, a retrotransposon from the fungal plant pathogen Fusarium
RT oxysporum."
RL Mol. Gen. Genet. 249:637-647(1995).
DR EMBL: L34658; AA88790.1;
DR InterPro: IPR001878; Znf_CCHC.1.
DR Pfam: PF00098; Zf-CCHC.1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_C2HC.1.
KW Polyprotein; Zinc-finger.
SQ SEQUENCE 853 AA; 99422 MW; 1AEDD18CFBBA5B06 CRC64;

Query Match
Best Local Similarity 15.9%; Score 121; DB 3; Length 853;
Matches 29; Conservative 12; Mismatches 31; Indels 12; Gaps 2;

OY 47 LPPAARRMNP1PPEPDGDTDRLEPFIYQTCMFVDENTSNDALKVFLIT 101
DB 100 LSAANNGRDPSEVLKPSPEYFDGTPSKLPFLTQSAFITYYPQFRDSAKVMYMG 159
OY 102 RLTPALQWVPIRKESPLNDY 125
DB 160 RLKTAQNF-----QPIINDY 176

RESULT 10
OY 096V43 PRELIMINARY; PRT; 1252 AA.
ID 096V43;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Aspergillus flavus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL 19997; TRANSPOSON-AFRTL-1;
RX MEDLINE-95314240; PubMed-7793909;
RA McAlpin C.E., Mannarelli B.;
RT "Construction and characterization of a DNA probe for distinguishing
RT strains of Aspergillus flavus."

```

```

RL Appl. Environ. Microbiol. 61:1068-1072(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL 19997: TRANSPOSON-AFRTL-1.
RA Okubata P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT "Airtl-1, a retrotransposon-like element in the aflatoxin-producing
RT fungus Aspergillus flavus."
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF362957; AAL26311.1; -
FT NON_TER 1252 1252
SQ SEQUENCE 1252 AA; 144752 MW; DEF0C5DD8B6C4B8 CRC64;

Query Match 13.8%; Score 105.5; DB 3; Length 1252;
Best Local Similarity 25.4%; Pred. No. 0.01;
Matches 36; Conservative 15; Mismatches 56; Indels 35; Gaps 3;

OY 11 RRLSADPATORNS---AAGTMDGROVQMLKALLAGPLRPAARMRNIPPEPTEGCD 67
DB 66 RIRNDOODHIAQIDAOVGASAPKDAIGKVKLPRA-----EPPDGT 105
OY 68 TURLPEFIVQCSYMFVDENTFSNDALKYTFILTRLGPALOWVPIYRK----- 117
DB 106 RSKDAFLQMMKHIANKNLIDEADKIFITSLRGAAAMNFEPIREYEVVDNWS 165
OY 118 --ESPFLANDYRGFLAEKRVFG 137
DB 166 NTRRELFTDSGNLRKHLERFEG 187

RESULT 11
O9HFY8 PRELIMINARY; PRT; 837 AA.
AC O9HFY8:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN.
GN GAG.
OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella
OS cingulata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes Incertae sedis; Phyllosthorales; Phyllosthoraceae;
OC Glomerella.
OX NCBI_TaxID=5457;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-RETROTRANSPOSON CGRET;
RA Zhu P., Oudemans P.V.;
RT "A long terminal repeat retrotransposon Cgret from the phytopathogenic
RT fungus Colletotrichum gloeosporioides on cranberry."
RL Curr. Genet. 0:0-0(2000).
DR EMBL: AF264028; AAG24791.1; -
DR InterPro: IPR001878; Znf_CCHC;
DR Pfam: PF00098; Zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.
SQ SEQUENCE 837 AA; 97738 MW; EF1D4BC70FD55003 CRC64;

Query Match 13.2%; Score 101; DB 3; Length 837;
Best Local Similarity 29.5%; Pred. No. 0.019;
Matches 36; Conservative 15; Mismatches 59; Indels 12; Gaps 3;

OY 25 SAARGTMDGROVQMLKALL--AGPLRPAAR-----RMRNPIPEPTEGCDTDLPEF 74
DB 66 NAAALKEIQARAREAGLKEVATLRAANVISAPEGREKLTAKTPTFTGCTGQLKGH 125
OY 75 IVQCSYMFVDENTFSNDALKYTFILTRLGPALOWVPIYRKESPLNDYRGFLAEKR 134
DB 126 LVQIRYQAFHMGTFQDTEFVHAATFLRGALAWPEPL--QOEWLNDPYEKYSOEVRN 183
OY 135 VF 136
DB 184 IF 185

```

```

RESULT 12
O9EQ11 PRELIMINARY; PRT; 231 AA.
AC O9EQ11:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6;
RX MEDLINE-21105984; PubMed-11158386;
RA Volfe J.-N., Koerting C., Scharl M.;
RT "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they
RT evolve into new cellular functions?";
RL Mol. Biol. Evol. 18:266-270(2001).
DR EMBL: AF302691; AAG39979.1; -
DR InterPro: IPR001878; Znf_CCHC;
DR Pfam: PF00098; Zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 231 AA; 26171 MW; DEAB2AZB624F3974 CRC64;

Query Match 12.7%; Score 97; DB 1; Length 231;
Best Local Similarity 30.9%; Pred. No. 0.011;
Matches 21; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

OY 69 DRLPEFIVQCSYMFVDENTFSNDALKYTFILTRLGPALOWVPIYRKESPLNDYRGF 128
DB 1 DMLGPFYQCQLFEMKSTROFSDYDRVCFYTSMLIGRAAWATAKLOCTYLMHNATF 60
OY 129 LAEMKRVF 136
DB 61 MMELEKHYF 68

RESULT 13
O9UVC2 PRELIMINARY; PRT; 639 AA.
AC O9UVC2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPEPTIDE (PUTATIVE GAG HOMOLOGUE).
GN GAG.
OS Cladosporium fulvum (Fulvia fulva).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetochytrionycetes Incertae Mycosphaerellaceae;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=5499;
RN [1]
RP SEQUENCE FROM N.A.
RC Simpson M.L., Butler M., Poulter R.T.M.;
RT "Functions of the integrase of retrotransposons: the integrase of the
RT Cfr-1 element from Cladosporium fulvum."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-CFR-1 LTR-RETROTRANSPOSON;
RX MEDLINE-92318885;
RA Mchale M.T., Roberts I.N., Noble S.M., Beaumont C., Whitehead M.P.,
RA Seth D., Oliver R.P.;
RT "Cfr-1, an LTR-retrotransposon in Cladosporium fulvum, a fungal
RT pathogen of tomato."
RL Mol. Gen. Genet. 233:337-347(1992).
DR EMBL: AF051915; AAF21677.1; -

```

DR EMBL: Z11866; CAA77890.1; -.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00098; Zf_CCHC; 1.
 DR SMART: SM00343; Znf_C2HC; 1.
 KW Polyprotein; zinc-finger.
 SO SEQUENCE 639 AA; 73649 MW; 4D3BA4B19AE0753A CRC64;

Query Match 11.1%; Score 85; DB 3; Length 639;
 Best Local Similarity 30.2%; Pred. No. 0.83;
 Matches 26; Conservative 8; Mismatches 40; Indels 12; Gaps 2;

OY 61 PETFGDTRPERIVGTCSTNFEDENTFSNDALKVFLTRLGPAQWVPIRK--- 117
 DB 19 PEFTGDRVFTWVSDMDVLFNSMT---ENLKPITFATFELGRADHWKPFPLKYL 75
 OY 118 -----ESPLNDYRGFLAEKRVFG 137
 DB 76 SNGEDNADGVFKSYNHLKHNKSVFG 101

RESULT 14

ID 09UVD9 PRELIMINARY; PRT; 406 AA.
 AC 09UVD9;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE GAG.
 GZ GAG.
 OS Alternaria alternata.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
 OX NCBI_Taxid=5599;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-15A; TRANSPOSON-LTR-RETROTRANSPOSON REAL;
 RA MEDLINE-20309324; PubMed-10852484;
 RX Kaneko I., Tanaka A., Tsuge T.;
 RT "REAL, an LTR-retrotransposon from the plant pathogenic fungus
 RT Alternaria alternata.";
 RL Mol. Gen. Genet. 263:625-634(2000).
 DR EMBL: AB025309; BAB89271.1; -.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR001993; Mitoch_carrier.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00098; Zf_CCHC; 1.
 DR SMART: SM00343; Znf_C2HC; 1.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
 DR Zinc-finger.
 SO SEQUENCE 406 AA; 46234 MW; F792231FC75CC8CF CRC64;

Query Match 11.0%; Score 84; DB 3; Length 406;
 Best Local Similarity 24.1%; Pred. No. 0.61;
 Matches 26; Conservative 19; Mismatches 61; Indels 2; Gaps 1;

OY 8 MRFRRRLSADPHATQRNSAEAGTMDGRVQLAKALLGPLPPAARWRNPPIPPETFGCD 67
 DB 89 MRLLSPASRLAEALERSAGESSSLGANTTPLSTPVAPAVR--KNKFPDEPRFDGT 146
 OY 68 TDRLPEFIVGTCSTNFEDENTFSNDALKVFLTRLGPAQWVPIYI 115
 DB 147 RGNPFGWTFCEGKLEYDCAMFTEADARVRYLSTRKDKANOVLLPWV 194

RESULT 15

ID 09SKK1 PRELIMINARY; PRT; 400 AA.
 AC 09SKK1;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE THEORETICAL 46.6 KDA PROTEIN.

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca;
 OX NCBI_Taxid=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CEREBELLUM CORTEX;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB060816; BAB46855.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 400 AA; 46612 MW; 911D7B5ADDOEB56 CRC64;

Query Match 10.9%; Score 83; DB 6; Length 400;
 Best Local Similarity 37.3%; Pred. No. 0.77;
 Matches 19; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

OY 86 ENTFSNDALKVFLTRLGPAQWVPIRKESPLNDYRGFLAEKRVF 136
 DB 11 EDHFGGAEVAVFLISFTGAEKDAISVTOGCSPLNANFPFLDEIRKF 61

Search completed: October 11, 2002, 02:43:08
 Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:39:14 ; Search time 31 Seconds

(without alignments)
515.956 Million cell updates/sec

Title: US-09-673-400A-38

Sequence: 1 ARAPTLDMRFRRLSADPHA.....YRGFLAEKRVGHEDEDF 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq_032802:*

1: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*

10: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	763	100.0	144	20	AAV59927
2	566	74.2	113	22	AAAB60475
3	549.5	72.0	142	22	ABG12205
4	350	45.9	106	21	AAAG04029
5	335	43.9	126	20	AAV25512
6	335	43.9	146	20	AAV26051
7	335	43.9	146	20	AAAB60492
8	328	43.0	66	21	AAAG02068
9	316	41.4	110	20	AAV26054
10	310.5	40.7	129	22	ABG12203
11	297	38.9	80	20	AAV26056

12	285	37.4	67	21	AAAG02067
13	227	29.8	42	22	ABG12204
14	201	26.3	65	20	AAV26058
15	197	25.8	184	21	AAV94673
16	194	25.4	168	21	AAV94674
17	183	24.0	74	20	AAV26053
18	182	23.9	51	20	AAV26057
19	164	21.5	44	20	AAV26055
20	157	20.6	79	21	AAV94690
21	157	20.6	98	21	AAV94691
22	154	20.2	65	21	AAV94689
23	139.5	18.3	1607	22	ABG15099
24	137	18.0	342	21	AAAB42148
25	134	17.6	57	21	AAV94693
26	134	17.6	76	21	AAV94694
27	131	17.2	43	21	AAV94692
28	85	11.1	45	20	AAV26052
29	76.5	10.0	676	22	ABG15498
30	76	10.0	41	21	AAV94688
31	71.5	9.4	397	22	AAAG91310
32	70.5	9.2	202	22	ABAB50194
33	70.5	9.2	202	22	AAAB92808
34	70.5	9.2	202	22	AAAB92808
35	70.5	9.2	243	21	AAAB43610
36	70.5	9.2	245	22	AAAG75627
37	70.5	9.2	688	21	AAAB07559
38	70.5	9.2	2386	22	ABG04328
39	70	9.2	358	22	AAAB79919
40	70	9.2	358	22	AAAB80051
41	70	9.2	363	22	AAAG92156
42	69	9.0	77	21	AAAB32065
43	69	9.0	77	21	AAAB32066
44	69	9.0	272	21	AAAG58307
45	69	9.0	348	22	AAV28026

ALIGNMENTS

RESULT 1
ID AAV59927 standard: Protein; 144 AA.
AAV59927;
28-JAN-2000 (first entry)
Human myometrium tumour EST encoded protein 7.
Myometrium; tumour; human; expressed sequence tag; EST; uterine myoma; treatment; carcinoma; cancer; gene therapy.
Homo sapiens.
DE19817947-A1.
28-OCT-1999.
17-APR-1998; 98DE-1017947.
17-APR-1998; 98DE-1017947.
(META-) METAGEN GBS GENOMFORSCHUNG MBH.
Rosenthal A, Specht T, Hinemann B, Schmitt A, Pilarsky C, Dahl E;
WPI; 1999-602380/52.
N-PSDB; AA241965.
New nucleic acid sequences expressed in uterine myoma, and derived polypeptides, for treatment of uterine carcinoma and identification of therapeutic agents -

PS Claim 23; Page 68; 86pp; German.
 XX
 CC (1) fragments and their encoding nucleic acids (II) which are highly
 CC expressed in human uterine myoma. (II) are used for recombinant
 CC expression of (I) and to isolate complete genes. (I) are used to
 CC identify agents suitable for treatment of uterine carcinoma, to directly
 CC treat this form of cancer (including expression from gene therapy
 CC vectors) and are used in a preparation for cancer treatment. (I) is also
 CC used for the generation of specific antibodies. (II) are identified by
 CC assembling ESTs (expressed sequence tags) from a particular tissue type
 CC before comparison of expression patterns. This allows a significantly
 CC longer fragment of the gene to be revealed and therefore reduces the
 CC number of failures associated with the fact that ESTs from different
 CC libraries may represent different parts of the same unknown gene,
 CC distorting the estimated frequency of occurrence in a particular tissue.
 CC AA59921-159940 represent protein fragments encoded by the human
 CC myometrium tumour cDNA library derived EST fragments represented in
 CC AA41950-241980.
 CC
 XX
 SO Sequence 144 AA;
 Query Match 100.0%; Score 763; DB 20; Length 144;
 Best Local Similarity 100.0%; Pred. No. 1.3e-83;
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ARAPLDNRFRRLSADHATORNSAEARCTMDGRVOLKALLAGPLRPAARWRNP 60
 Db 1 ARAPLDNRFRRLSADHATORNSAEARCTMDGRVOLKALLAGPLRPAARWRNP 60
 Oy 61 PETFDGDTDLRPEFIYOTCSYMFVDENTFSNDALKVTFLITRLTGPALOWVPIRKESP 120
 Db 61 PETFDGDTDLRPEFIYOTCSYMFVDENTFSNDALKVTFLITRLTGPALOWVPIRKESP 120
 Oy 121 LLNDYRGFLAEKRRVFGWEDEDF 144
 Db 121 LLNDYRGFLAEKRRVFGWEDEDF 144
 RESULT 2
 AAB60475
 ID AAB60475 standard; Protein; 113 AA.
 XX
 AC AAB60475;
 XX
 DT 24-APR-2001 (first entry)
 DE Human cell cycle and proliferation protein CCYPR-23, SEQ ID NO:23.
 XX
 KM Cell cycle and proliferation protein; CCYPR; human; agonist;
 KM antagonist; gene therapy; detection; gene therapy;
 KM transgenic animal disease model; immune disorder;
 KM developmental disorder; cell signalling disorder;
 KM cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KM arteriosclerosis; asthma; allergy; diabetes mellitus;
 KM menstrual cycle disorder; bacterial infection.
 XX
 OS Homo sapiens.
 XX
 PN WO200107471-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000WO-US19948.
 XX
 PR 21-JUL-1999; 99US-0145075.
 PR 08-SEP-1999; 99US-0153129.
 PR 10-NOV-1999; 99US-0164647.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O,
 PI Azimzai Y, Yang J, Lu DM, Baughn MR, Patterson C, Shah P;

XX
 DR WPI; 2001-112727/12.
 DR N-PSDB; AAF59612.
 XX
 PT Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signalling disorders and cell proliferative disorders including cancer
 PS
 XX Claim 1; Page 131; 205pp; English.
 XX
 CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and
 CC proliferation proteins (CCYPR), which are encoded by AAF5950-AAF59643.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC Diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders, and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC diabetes mellitus, disorders of the menstrual cycle and infections
 CC caused by bacteria.
 CC
 XX
 SO Sequence 113 AA;
 Query Match 74.2%; Score 566; DB 22; Length 113;
 Best Local Similarity 93.8%; Pred. No. 4.3e-60;
 Matches 106; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Oy 32 MDGRVOLKALLAGPLRPAARWRNPPEFDGDTDLRPEFIYOTCSYMFVDENTFSN 91
 Db 1 MDGRVOLKALLAGPLRPAARWRNPPEFDGDTDLRPEFIYOTCSYMFVDENTFSN 91
 Oy 92 DALKVTFLITRLTGPALOWVPIRKESPLNDYRGFLAEKRRVFGWEDEDF 144
 Db 61 DALKVTFLITRLTGPALOWVPIRKESPLNDYRGFLAEKRRVFGWEDEDF 113
 RESULT 3
 ABG12205
 ID ABG12205 standard; Protein; 142 AA.
 XX
 AC ABG12205;
 XX
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #12196.
 XX
 KM Human; Chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.

PI	Dmanac RT, Liu C, Tang YT:
XX	
DR	WPI; 2001-639362/73.
XX	N-PSDB; AAS76392.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	ClaIm 20; SEQ ID No 42564; 103pp: English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	spectification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 142 AA;
Query Match	72.0%: Score 549.5; DB 22; Length 142;
Best Local Similarity	74.8%; Pred. No. 5.8e-58;
Matches 107; Conservative 12; Mismatches 19; Indels 5; Gaps 2	
OY	6 LDMRRRLSLADPHNQRNSAARGMGR----VOLMKALAGPLRPARMRNPPIP 61
Db	1 ISLRILDGIAGCTHASATEASDR-SEDLKMTVCALOIALALPRLPRTRRRNPIPP 59
OY	62 ETFDDDTLRLPEFIYOTCSYMFVDENTESNDALKVTFLITRLTGAPALQWVIPYIKRESPL 121
Db	60 ETFDDDTLRLPEFIYOTCSYMYVDENTFESSDALKVTFLITRLTGAPALQWVIPYIKRESPL 119
OY	122 LNDYRGFLAEKRVFGWEDED 144
Db	120 LNDYRGFLAEKRVFGWEDED 142
RESULT 4	
AAG04029	
ID	AAG04029 standard; Protein; 106 AA.
AC	AAG04029;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein, SEQ ID NO: 8110.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM	gene therapy; chromosome mapping.
OS	Homo sapiens.
XX	
PN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	

[illegible]

XX	Cell cycle and proliferation protein; CCYPR; human; agonist;
KW	antagonist; gene therapy; detection; gene therapy;
KW	transgenic animal disease model; immune disorder;
KW	developmental disorder; cell signalling disorder;
KW	cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW	arteriosclerosis; asthma; allergy; diabetes mellitus;
KW	menstrual cycle disorder; bacterial infection.
XX	
OS	Homo sapiens.
XX	
PN	W0200107471-A2.
XX	
PD	01-FEB-2001.
XX	
PF	21-JUL-2000; 2000MO-US19948.
XX	
PR	21-JUL-1999; 99US-0145075.
PR	08-SEP-1999; 99US-0153129.
PR	10-NOV-1999; 99US-0164647.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
PI	Hillman JL, Lal P, Tang YT, Yee H, Au-Young J, Bandman O;
PI	Azmiel Y, Yang J, Lu DM, Baughn MR, Patterson C, Shah P;
XX	
DR	WPI: 2001-112727/12.
XX	
PT	Human cell cycle and proliferation proteins and polynucleotides are
PT	used to treat, diagnose and prevent immune, developmental and cell
PT	signalling disorders and cell proliferative disorders including cancer -
XX	
PS	Disclosure: Page 150; 205pp; English.
XX	
CC	Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC	proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC	CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC	associated with decreased expression of functional CCYPR, while CCYPR
CC	antagonists are used to treat diseases or conditions associated with
CC	overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC	to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC	radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC	compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC	that specifically bind to CCYPR, and in drug screening methods to
CC	identify compounds that modulate the activity of CCYPR. CCYPR
CC	nucleotides can be used to generate transgenic animal models of human
CC	disease, and can be used in gene therapy in target cells with genetic
CC	abnormalities with respect to the expression of CCYPR for the
CC	treatment or prevention of a disorder associated with CCYPR.
CC	Diseases which can be diagnosed, treated and prevented using CCYPR
CC	proteins, nucleic acids, agonists or antagonists include immune,
CC	developmental and cell signalling disorders, and cell proliferative
CC	disorders including cancer. Specific examples of these disorders
CC	include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC	diabetes mellitus, disorders of the menstrual cycle and infections
CC	caused by bacteria.
XX	
SO	Sequence 146 AA:
XX	
Query Match	43.9%; Score 335; DB 22; Length 146;
Best Local Similarity	59.6%; Pred. No. 3.6e-32;
Matches 62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;	
QY	40 KALLAGPLPARRRRNRNPPEPTDGDTRLRPEPTVQCSWFYDENFFSDALKVPL 99
DB	37 RASTLRQVPPSP---CPVPPEPTENGESSRLPEFTVQASVTLNENRCDAMKVAFL 92
QY	100 ITRLTGPALOWIPIYTRKESPLNDYRGRLAEMKRYFGREEDD 143
DB	93 ISLTGGAEMVVPYIENDSPILGDIKRALFLDKKQCFGMDDED 136

RESULT 8	
AAG02068	
ID	AAG02068 standard; Protein: 66 AA.
XX	
AC	AAG02068:
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein, SEQ ID NO: 6149.
XX	
KM	Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
RW	gene therapy; chromosome mapping.
XX	
OS	Homo sapiens.
XX	
FN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PE	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST) GENSET.
XX	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
DR	WPI: 2000-500381/45.
DR	N-PSDB: AAC02074.
XX	
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX	
PS	Claim 13; SEQ ID 6149; 71pp + CD-ROM; English.
XX	
CC	The present sequence is a polypeptide encoded by one of a large number
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30
CC	different tissues. EST sequences usually correspond mainly to the 3'
CC	untranslated region (UTR) of the mRNA because they are often obtained
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC	those cases where longer cDNA sequences have been obtained, the full 5'
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC	ends and can therefore be used to obtain full length cDNAs and genomic
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC	chromosome mapping procedures. They are used to obtain upstream
CC	regulatory sequences and to design expression and secretion vectors.
XX	
SQ	Sequence 66 AA:
Query Match	43.0%; Score 328; DB 21; Length 66;
Best Local Similarity	95.5%; Pred. No. 9, 2e-32;
Matches 63; Conservative 0; Mismatches 3; Indels 0; Gaps 0	
QY	32 MDGRVQLMKALLAGPIRAPARRWRNPDPPEPTFDGDTDRLPDEFIVTQCYMEVDENTFSN 91
DB	1 MKGRVQLMKALLAXPIRAPARRWRNPDPPEPTFDGDTDRLPDEFIVTQSSMEVDENTFSN 60
QY	92 DALKVT 97
DB	61 DALKVT 66
RESULT 9	
AAZ26054	
ID	AAZ26054 standard; Protein: 110 AA.
XX	
AC	AAZ26054:
XX	
DT	28-SEP-1999 (first entry)
XX	

XX 17-JUN-1999.
 PD 10-DEC-1998; 98WO-US26273.
 XX 10-DEC-1997; 97US-0987926.
 PR 10-DEC-1997; 97US-0987926.
 XX (ZYMO) ZYMOGENETICS INC.
 PA (ZYMO) ZYMOGENETICS INC.
 XX Conklin DC, Lok S, Parrish J;
 PI WPI; 1999-385572/32.
 XX Mammalian alpha helical protein-1, designated zalpal
 PT Claim 10; Page 72; 73pp; English.
 XX The present sequence is an epitope-bearing protein derived from
 CC human alpha helical protein-1, zalpal. The protein comprises
 CC helices B, C and D of mature zalpal. It is used to raise specific
 CC antibodies which can be used for detection and purification of zalpal.
 CC The zalpal protein may be useful in the treatment of Fragile-X
 CC syndrome and abnormal proliferative disorders e.g. cancer. It can also be
 CC used for the growth, differentiation, maintenance and survival of
 CC connective tissues, particularly cardiovascular and epidermal systems
 CC and in imparting cosmetic improvements to normal connective tissues such
 CC as enhancement of skin tone and elasticity.
 XX Sequence 80 AA:
 SQ
 Query Match 38.9%; Score 297; DB 20; Length 80;
 Best Local Similarity 66.2%; Pred. No. 6.2e-28;
 Matches 53; Conservative 12; Mismatches 15; Indels 0; Gaps 0;
 QY 60 FPEFDCDTRLPEFIVQCSYMFVDETFESNDALKTFLITRLTGPLALOWIPIRKES 119
 Db 1 FPEFENGSSRLPEFIVQCSYMFVDETFESNDALKTFLITRLTGPLALOWIPIRKES 60
 QY 120 PLINDYRGFLAEKRVFGME 139
 Db 61 PILGDYRAFLDEMKOCFGMD 80

RESULT 12
 AAG02067
 ID AAG02067 standard; Protein: 67 AA.
 AC AAG02067;
 XX 06-OCT-2000 (first entry)
 DE Human secreted protein. SEQ ID NO: 6148.
 DE Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX Homo sapiens.
 OS Homo sapiens.
 XX EP1033401-A2.
 PN 06-SEP-2000.
 PD 21-FEB-2000; 2000EP-0200610.
 PF 26-FEB-1999; 99US-0122487.
 PR (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR N-PSDB; AAC02073.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 13; SEQ ID 6148; 71pp + CD-ROM; English.
 XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX Sequence 67 AA:
 SQ
 Query Match 37.4%; Score 285; DB 21; Length 67;
 Best Local Similarity 86.9%; Pred. No. 1.4e-26;
 Matches 53; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 32 MDGRVQLKALLAPLRPARRRNPFPETFDGDTDLPEFIVQCSYMFVDETFESN 91
 Db 1 MDGRVQLKALLAPLRPARRRNPFPETFDGDTDLPEFIVQCSYMFVDETFESN 60
 QY 92 D 92
 Db 61 E 61

RESULT 13
 ABG12204
 ID ABG12204 standard; Protein: 42 AA.
 AC ABG12204;
 XX 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #12195.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS76391.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID NO 42563; 103pp; English.
 XX

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 42 AA:
 SQ
 Query Match 29.8%; Score 227; DB 22; Length 42;
 Best Local Similarity 97.6%; Pred. No. 7.1e-20;
 Matches 41; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 103 LTGPALQWVPIPIRKESPLNDYRGFLAEKRVGMEDEDE 144
 Db 1 LTGPALQWVPIPIRKESPLNDYRGFLAEKRVGMEDEDE 42
 RESULT 14
 AAY26058
 ID AAY26058 standard; Protein: 65 AA.
 XX
 AC AAY26058;
 DT 28-SEP-1999 (first entry)
 XX
 DE Human Zalphal epitope-bearing protein fragment 7.
 XX
 KW Alpha helical protein-1; Fragile-X syndrome; cardiovascular system;
 KW connective tissue; abnormal proliferative disorder; cancer; skin tone;
 KW epidermal system; cosmetic improvement; skin tone; elasticity; Zalphal;
 KW epitope.
 XX
 OS Homo sapiens.
 XX
 PN WO9929720-A2.
 XX
 PD 17-JUN-1999.
 XX
 PP 10-DEC-1998; 98WO-US26273.
 XX
 PR 10-DEC-1997; 97US-0987926.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Lok S, Parrish J;
 XX
 DR WPI; 1999-385572/32.
 XX
 XX Mammalian alpha helical protein-1, designated Zalphal
 PS Claim 10; Pages 72-73; 73pp; English.
 CC The present sequence is an epitope-bearing protein derived from
 CC human alpha helical protein-1, Zalphal. It is used to raise specific
 CC antibodies which can be used for detection and purification of Zalphal.
 CC The Zalphal protein may be useful in the treatment of Fragile-X
 CC syndrome and abnormal proliferative disorders e.g. cancer. It can also be

CC used for the growth, differentiation, maintenance and survival of
 CC connective tissues, particularly cardiovascular and epidermal systems
 CC and in imparting cosmetic improvements to normal connective tissues such
 CC as enhancement of skin tone and elasticity.
 XX
 SQ Sequence 65 AA:
 Query Match 26.3%; Score 201; DB 20; Length 65;
 Best Local Similarity 63.6%; Pred. No. 1.7e-16;
 Matches 35; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
 Oy 89 FSNALRYTFLITRLTGPALQWVPIPIRKESPLNDYRGFLAEKRVGMEDEDE 143
 Db 1 FSNALRYTFLITRLTGPALQWVPIPIRKESPLNDYRGFLAEKRVGMEDEDE 55
 RESULT 15
 AAY94673
 ID AAY94673 standard; Protein: 184 AA.
 XX
 AC AAY94673;
 DT 01-DEC-2000 (first entry)
 XX
 DE Human zsi983 protein sequence.
 XX
 KW Alpha-helical protein; zsi983; cell growth; differentiation; cancer;
 KW proliferation; chromosome 22q13.1-q13.2; cytostatic; vulnery;
 KW degenerative condition; metastasis; wound healing.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Location/Qualifiers
 FT 1..16
 FT /note="Putative signal sequence"
 FT 17..184
 FT /label="zsi983"
 FT 38..54
 FT /note="SH3-binding domain"
 FT 57..71
 FT /label="Helix_A"
 FT 79..97
 FT /label="Helix_B"
 FT 76..81
 FT /note="Hydrophilic region"
 FT 90..95
 FT /note="Hydrophilic region"
 FT 113..121
 FT /label="Helix_C"
 FT 119..124
 FT /note="Hydrophilic region"
 FT 126..135
 FT /label="Helix_D"
 FT 144..154
 FT /label="Helix_E"
 FT 142..147
 FT /note="Hydrophilic region"
 FT 178..183
 FT /note="Hydrophilic region"
 XX
 PN WO200050594-A2.
 XX
 XX 31-AUG-2000.
 XX
 PD 25-FEB-2000; 2000WO-US04816.
 XX
 PF 26-FEB-1999; 99US-0259131.
 XX
 PR (ZYMO) ZYMOGENETICS INC.
 XX
 PA Presnell SR;
 XX
 DR WPI; 2000-572091/53.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:40:54 ; Search time 20 Seconds

(without alignments)
691.843 Million cell updates/sec

Title: US-09-673-400a-38

Perfect score: 763
Sequence: 1 ARAPTLDMRRRRRLSADPHA.....YRGFLAEKKRVFGMEDEDF 144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	15.9	853	2	S60178
2	104	13.6	457	2	T18347
3	85	11.1	639	2	S23569
4	78	10.2	223	2	JCS323
5	77.5	10.2	499	2	E87416
6	77	10.1	232	2	JE0163
7	74.5	9.8	464	2	T01324
8	74	9.7	414	2	F70778
9	73.5	9.6	589	2	G83758
10	73	9.6	853	2	T08162
11	71.5	9.4	485	2	T35628
12	71.5	9.4	550	2	T21194
13	70.5	9.2	180	2	D82550
14	70.5	9.2	443	2	S59771
15	70	9.2	1442	2	T42607
16	69.5	9.1	476	2	E83796
17	69.5	9.1	1609	2	T01797
18	69	9.0	348	1	R5HUL3
19	69	9.0	1487	1	EDBE1
20	68.5	9.0	348	2	T36890
21	68.5	9.0	471	2	B84412
22	68.5	9.0	586	2	AF1136
23	68.5	9.0	1515	1	S51863
24	68	8.9	250	2	A83310
25	68	8.9	1415	2	C86438
26	67.5	8.8	201	2	AB2216
27	67.5	8.8	316	2	S25843
28	67.5	8.8	971	2	S28833
29	67	8.8	729	2	F86308

30	66.5	8.7	269	2	G90261
31	66.5	8.7	325	2	G72020
32	66.5	8.7	337	2	A53041
33	66.5	8.7	430	2	F87112
34	66.5	8.7	441	2	A96759
35	66.5	8.7	769	2	H97033
36	66.5	8.7	914	1	JN0550
37	66.5	8.7	914	1	S07047
38	66.5	8.7	1140	2	F88349
39	66.5	8.7	1140	2	T20984
40	66	8.7	130	2	F95328
41	66	8.7	290	2	S41555
42	66	8.7	290	2	S69841
43	66	8.7	875	2	F70755
44	66	8.7	1058	2	T19282
45	66	8.7	1498	2	S53577

ALIGNMENTS

RESULT 1
S60178
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy
C:Species: Fusarium oxysporum
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C:Accession: S60178
R:Anaya, N.; Roncero, M.I.G.
Mol. Gen. Genet. 249, 637-647, 1995
A:Title: Skippy, a retrotransposon from the fungal plant pathogen Fusarium oxysporum
A:Reference number: S60178; MID:96132549
A:Accession: S60178
A:Molecule type: DNA
A:Residues: 1-853 <ANA>
A:Cross-references: EMBL:L34658; NID:9510695; PIDN:AAA8790.1; PID:9510696
A:Experimental source: retrotransposon skippy
C:Genetics:
A:Mobile element: retrotransposon skippy
C:Keywords: polyprotein

Query Match 15.9%; Score 121; DB 2; Length 853;
Best Local Similarity 34.5%; Pred. No. 0.00012;
Matches 29; Conservative 12; Mismatches 31; Indels 12; Gaps 2;

OY 47 LRPAARRMRPIPF-----DETDCGTDRLEPIVTCSTMEVDENTFSMDALKVFLIT 101
Db 100 LASAANGRDPEVLKSPPEYFGDTPSKPTFLTOSRAFTYYPNOFRNDSAKVMTMG 159
OY 102 RLTCPALQWVTPYRKESPLINDY 125
Db 160 RLTKTAQWF-----QPIINDY 176

RESULT 2
T18347
gag protein homolog, truncated - rice blast fungus magnaporthe gypsy retrotransposon
C:Species: Magnaporthe grisea (rice blast fungus)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T18347
R:Farman, M.L.; Tosa, Y.; Nitta, N.; Leong, S.A.
Mol. Gen. Genet. 251, 665-674, 1996
A:Title: MAGT, a retrotransposon in the genome of the rice blast fungus Magnaporthe
A:Reference number: Z18882; MUID:96335141
A:Accession: T18347
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-457 <FAR>
A:Cross-references: EMBL:L35053; NID:9522300; PID:9522301; PIDN:AAA33419.1
C:Genetics:
A:Gene: gag
A:Mobile element: magnaporthe gypsy retrotransposon

Query Match 13.6%; Score 104; DB 2; Length 457;

Best Local Similarity 27.0%; Pred. No. 0.0038;
Matches 24; Conservative 17; Mismatches 48; Indels 0; Gaps 0;

OY 49 PAKRRRNPIPPETFDGTDRLPEFIVOTCSMFEDENTFSNDALKTFLIRLGLPAL 108
DB 152 PASARISERLPPDKPTGARSRLRRPATQIRKMTSNKDPFNPESRLIYIAGRLSKAY 211

OY 109 QWVPIRKESPLINDRGFLAEKRVFG 137
DB 212 NLLPKMVGCTPQFGDTLLQYLEARF 240

RESULT 3

gag polyprotein homolog - fungus (Cladosporium fulvum) retrotransposon CFT-1
C:Species: Cladosporium fulvum

C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 21-Jul-2000
C:Accession: S23569; S19849

R:McHale, M.T.; Roberts, I.N.; Noble, S.M.; Beaumont, C.; Whitehead, M.P.; Seth, D.; Oik
Mol. Gen. Genet. 233, 337-347, 1992

A:Title: CFT-1: an LTR-retrotransposon in Cladosporium fulvum, a fungal pathogen of toma
A:Reference number: S23569; MID:92318865

A:Accession: S23569
A:Molecule type: DNA

A:Residues: 1-639 <MCCH>
A:Cross-references: EMBL:Z11666; NID:q2562; PIDN:CAA77890.1; PID:q2563

A:Note: the authors translated the codon AACC for residue 55 as Ala, CAA for residue 451
A:Genetics:

A:Mobile element: retrotransposon CFT-1
C:Superfamily: Cladosporium fulvum probable gag polyprotein
C:Keywords: polyprotein

Query Match 11.1%; Score 85; DB 2; Length 639;
Best Local Similarity 30.2%; Pred. No. 0.63;
Matches 26; Conservative 8; Mismatches 40; Indels 12; Gaps 2;

OY 61 PETFDSDTDLRLPEFIVOTCSYMEVDENTFSNDALKTFLIRLGLPALQWVPIRK-- 117
DB 19 PEPFYGDGVAFDTWVSQMDYFLFNSMT---ENLKPIFATFLTGRAQHWWKPLRKYL 75

OY 118 -----ESPLINDRGFLAEKRVFG 137
DB 76 SNGEDNADGVFKSYNHLKHAKEVFG 101

RESULT 4

JC6323
endonuclease (EC 3.1.21.-) - Bacillus globigii

C:Species: Bacillus globigii
C:Date: 14-May-1998 #sequence_revision 29-May-1998 #text_change 24-Nov-1999
C:Accession: JC6323

R:Anton, B.P.; Heiter, D.F.; Benner, J.S.; Hess, E.J.; Greenough, L.; Moran, L.S.; Slack
Gene 187, 19-27, 1997

A:Title: Cloning and characterization of the BglII restriction-modification system reve
A:Reference number: JC6323; MID:9725792

A:Accession: JC6323
A:Molecule type: DNA

A:Residues: 1-223 <NNT>
A:Cross-references: GB:U49842; NID:q1293617; PIDN:AAC45060.1; PID:q1293619

A:Experimental source: R08562
C:Genetics:

A:Gene: bglIIR
C:Superfamily: Bacillus globigii endonuclease

C:Keywords: hydrolase

Query Match 10.2%; Score 78; DB 2; Length 223;
Best Local Similarity 28.7%; Pred. No. 0.97;
Matches 33; Conservative 16; Mismatches 36; Indels 30; Gaps 7;

OY 52 RRANPIPPETFD-----GTDRLPEFIVOTCSYMEVDENT-----FSN 91
DB 64 KHKNNIPIRKRFPLGIDIDGKRDLYE--VQSNIPFLNNTVRSELFHKSMDIDE 121

OY 92 DALKTFLIRLGLPALQWVPIRKESPL--LNDYRGFLAEKRVFGMEED 141
DB 122 EGMKVALITRK--GHMFPASSSLYYEAOQNOLNSLAIFYVDFVPI-RLVGLIED 173

RESULT 5

E87416
monooxygenase, flavin-binding family [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87416

R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.;
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MID:21173698; PMID:11259647

A:Accession: E87416
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-499 <STO>

A:Cross-references: GB:AE005673; NID:q13422697; PIDN:AAK23329.1; GSPDB:GN00148
C:Genetics:

A:Gene: CCI348

Query Match 10.2%; Score 77.5; DB 2; Length 499;
Best Local Similarity 22.4%; Pred. No. 3;
Matches 46; Conservative 21; Mismatches 63; Indels 75; Gaps 10;

OY 11 RRLSADPHATORSALRGTGMDGRVOLMKRL-----LAQPLPAAARRNPIRPF 61
DB 252 RRRVLFQHEFTRAIEESDTV--KAELEGVKMFLEDFQVAAHFTPRVPRHORAFLV 309

OY 62 ETPFGD-----FDRLPEF-----IVOTCS--YMEVDENTF 89
DB 310 P--DGDLPFGVASKASAVYDEICFTKGLLKSGETLDADYITATGFLSVLGDI 367

OY 90 SND-----ALKVFLIRLGLPALQWVPIRKESPLNDYNG-----FLAEK---- 133
DB 368 EIDGQPLDFAKTVYRGMEFTGVPNLVWFGYFRASWTLRADLIGDFVCRLAHMERKGA 427

OY 134 -----RVFGMEDEDF 144
DB 428 KQVEALRPEDKDKMTIGMDIDEDF 452

RESULT 6

JE0163
myelin expression factor-3 - mouse

N:Alternate names: MyEF-3
C:Species: Mus musculus (house mouse)

C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C:Accession: JE0163

R:Stephewski, A.; Krynka, B.; Treliakova, A.; Haas, S.; Khalili, K.; Amin, S.
Biochem. Biophys. Res. Commun. 243, 295-301, 1998

A:Title: MyEF-3, a developmentally controlled brain-derived nuclear protein which s
A:Reference number: JE0163; MID:98139908

A:Accession: JE0163
A:Molecule type: mRNA

A:Residues: 1-232 <STP>
A:Experimental source: brain

C:Keywords: phosphoprotein
F:40-60/Domain: transmembrane #status predicted <TM>
F:31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pre

F:31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #st

Query Match 10.1%; Score 77; DB 2; Length 232;
Best Local Similarity 38.3%; Pred. No. 1.3;
Matches 18; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

OY 61 PETFDGTDRLPEFIVOTCSYMEVDENTFSNDALKTFLIRLGLPAL 107
DB 7 PEKFDGNFDMIGPFMYOCOLFMKSTROPSVDRIVCVTSLIGRA 53

RESULT 7

T01324

hypothetical protein III_4H10.1 - maize

C:Species: Zea mays (maize)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 29-Oct-1999

C:Accession: T01324

R:Liaca, V.; Messing, J.

A:Description: Structure and organization of the 22-kDa alpha zeln gene cluster in Zea m

A:Reference number: Z14294

A:Accession: T01324

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-464 <L1A>

A:Cross-references: EMBL:AF031569; NID:g2832242; PIDN:AC01574.1; PID:g2832244

Query Match

Best Local Similarity 20.8%; Score 74.5; DB 2; Length 464;

Matches 32; Conservative 17; Mismatches 60; Indels 45; Gaps 5;

Db 12 RLRSADPHATQRNSAARGTMDGRVQLMKALLAGPLRPAARMNPPI----- 59

Db 139 RRAGEARVSLERACERONIEGRNLDQDAEVVPOAPMGTRSOAGVPLACGCAALVDH 198

Qy 60 -----FPETFGDGTDRLPETIVQTCYSMEVDNTEFSDALKYFLITRLTGP 106

Db 199 LRASMPKRPRLPEKVDGTSN--PSEFLOV--YVATITAGCNTAMATYFHVAGSGP 254

Qy 107 ALQWVPIYIRKESPLNDYRGFLAEKRVFGWEE 140

Db 255 ARTW-----LNMISPG-----SIYSWEE 272

RESULT 8

F70778

hypothetical protein RV2242 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70778

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

A:Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

A:Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

A:Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: F70778

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-414 <COL>

A:Cross-references: GB:Z70692; GB:AL123456; NID:g3261567; PIDN:CAA94663.1; PID:e235181;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV2242

Query Match

Best Local Similarity 25.8%; Score 74; DB 2; Length 414;

Matches 24; Conservative 17; Mismatches 32; Indels 20; Gaps 4;

Qy 8 MRFRRLSADPHATQRNSAARGTMDGRVQLMKALLAGPLR-----AARRW----- 54

Db 141 LKYSKDLAFTATATVADAAEARGTMDSEASVDAVARGDTPGLLSRAALNDTAP 200

Qy 55 -----RNPFP--ETFGDTRLPETIVQTC 80

Db 201 ATVLVGTAPGPGNSGDSERASQVDRTPAA 233

RESULT 9

G83758

oligoendopeptidase F BH0871 [Imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: G83758

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F

A:Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodure

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: G83758

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-589 <STO>

A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04590.1; GSI

C:Genetics:

A:Gene: BH0871

Query Match

Best Local Similarity 24.2%; Score 73.5; DB 2; Length 589;

Matches 36; Conservative 17; Mismatches 49; Indels 47; Gaps 7;

Qy 5 TLDKRRRLSADPHATQRNSAARGTMDGRVQLMKALLAGPLRPAARMNPPI-----R 55

Db 129 TFSINERQAARLAPQERLVAQAVD-----YANSEVYQAVGR 172

Qy 56 NPFPETFGDGTDRLPETIVQTCYSMEVDNTEFSDALKYFLITRLTGPALQWVPIYI 115

Db 173 MEIPFE--DETRRLSGOL-----SNRLSSKRRYKRSKAYAE--W----- 214

Qy 116 RKESPL-----LNDYRGFLAEKRVFGWEE 140

Db 215 KKETPLFASTLNHLAGFLNLYRAGWMD 243

RESULT 10

T08162

amylopullulanase (EC 3.2.1.-) - Pyrococcus furiosus

C:Species: Pyrococcus furiosus

C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999

C:Accession: T08162

R:Dong, G.; Vieille, C.; Zeikus, J.G.

A:Appl. Environ. Microbiol. 63, 3577-3584, 1997

A:Title: Cloning, sequencing, and expression of the gene encoding amylopullulanase

A:Reference number: Z16389; MUID:97438521

A:Accession: T08162

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-853 <DON>

A:Cross-references: EMBL:AF016588; NID:g2435436; PIDN:AAB71229.1; PID:g2435438

C:Genetics:

A:Gene: apu

C:Function:

A:Description: hydrolyzes alpha-1,6 and alpha-1,4 linkages in starch-related pol;

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match

Best Local Similarity 22.7%; Score 73; DB 2; Length 853;

Matches 37; Conservative 16; Mismatches 52; Indels 58; Gaps 7;

Qy 21 TQRNSAARGTMDGRVQLMKALLAGPLRPAARMNPFPETFGDGTDRLE----- 73

Db 145 TDENGNPIRDFWDRTELKDKMLA-----AKQKYN-LPLEBQKAVVNEFTBODYDLA 198

Qy 74 -----PIVOT-----CSYMEVDNTEFSDALKYFLITRL----- 103

Db 199 VLFNLAWIDYNYLIISPELKALYDKVDEGYSREDK-TVLYHQMWLNNTEFEHEKINL 257

Qy 104 ---TGPALQWVPIYIRKESPLNDYRGFLAEKRVFGWEE 143

Db 258 LLGNQVEVTVVYAHPIGPIILND-----FGWSDFD 209

RESULT 11

T35628

probable penicillin-binding protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35628

R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, June 1999

A:Reference number: 221584

A:Accession: T35628

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-485 <SEP>

A:Cross-references: EMBL:AL079336; PIDN:CA645623.1; GSPDB:GN00070; SCOEDB:SC669.32

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC669.32

Query Match 9.4%; Score 71.5; DB 2; Length 485;

Best Local Similarity 26.6%; Pred. No. 13;

Matches 33; Conservative 15; Mismatches 43; Indels 33; Gaps 5;

12 RLISADPHATQNSAEAR---GTMDGRVOLMKALLAGPLRPAARRMRNPIPEPFDDG 67

193 KILNADPDKPALNRAVSKRTPPGSTFKVYTAALALPAGVIRDLDAFTSPDPF--TLPGT 250

68 TDRLP-----EFIVTQSYMF---VDENT-----FSNDALKYTFPI 100

251 RRLTFNEADGCRNALSREALFEMSCNTVFPAKLGVDGVCVGMTAMAEAFNFDGLRVFPV 310

101 TRLT 104

311 ARST 314

RESULT 12

T21194

hypothetical protein F21D5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T21194

R:Berts, M.

submitted to the EMBL Data Library, September 1995

A:Reference number: 219389

A:Accession: T21194

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-550 <WILD>

A:Cross-references: EMBL:Z54271; PIDN:CAA91032.1; GSPDB:GN00022; CESP:F21D5.1

A:Experimental source: Clone F21D5

C:Genetics:

A:Gene: CESP:F21D5.1

A:Map position: 4

A:Intons: 33/1; 97/3; 213/2; 329/3; 409/2; 512/2

C:Superfamily: Caenorhabditis elegans hypothetical protein F21D5.1

Query Match

Best Local Similarity 29.7%; Pred. No. 15;

Matches 38; Conservative 13; Mismatches 40; Indels 37; Gaps 8;

9 RFRRLSADPHATQ---RMSAE--RGTMGRVOLMKALLAGPLRPAARRMRNPIPEPE 62

230 RFRNLRLPSSLLVEFRNESELNCGADPVKISOKLPAN-FSPTALE-----PKCA 283

63 TFDGDTDRLEPEFIVTQCSYFVDENTFSNDAL-----KTFLLTRLTGPALOWVPIYIRK 117

284 SFGDADRLMTYFRAK-----ASENSSDNAELFDGKXIVLTI-----VYIIRE 326

118 ESPLANDY 125

327 Q---LMDY 331

RESULT 13

D82550

conserved hypothetical protein XF2493 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: D82550

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide S

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82517; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 <SIM>

A:Cross-references: GB:AE004057; GB:AE003849; NID:g9107690; PIDN:AAF65291.1; GSPDB:

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alverenga,

Bitones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.O.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.;

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.;

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Marti

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki,

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeira,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanceli, R.V.; Se

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Si

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2493

C:Superfamily: Pseudomonas aeruginosa hypothetical protein PRF10

Query Match 9.2%; Score 70.5; DB 2; Length 180;

Best Local Similarity 26.8%; Pred. No. 4.7;

Matches 33; Conservative 16; Mismatches 39; Indels 35; Gaps 7;

4 PTLDMFRRL--SADP--HATQNSAEARGTMDGRVOLMKALLAGPLRPAARRMRNPI 58

68 PAL-LRFQRLVIGHRDPFTTHQRLAAALGTW-----LYORRRRG-1 109

59 P-----FPETPGDTRLEPEFIVTQCSYFVDENTFSNDALKVTPL---ITRLTGP 106

110 PTPDCRVIALYPDEFNAELDRVYVWQDPVFLGNSAMTHAGTLPHTLMLSTFQTG 169

107 ALQ 109

170 SFQ 172

RESULT 14

S59771

hypothetical protein YPR106w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein P8283.9

C:Species: Saccharomyces cerevisiae

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 06-Feb-1998

C:Accession: S59771

R:Nelson, J.

submitted to the EMBL Data Library, July 1995

A:Description: The sequence of S. cerevisiae cosmid 8283.

A:Reference number: S59764

A:Accession: S59771

A:Molecule type: DNA

A:Residues: 1-443 <NLS>

A:Cross-references: EMBL:U02445; NID:g914969; PID:g914977; MIPS:YPR106w

C:Genetics:

A:Gene: SGD:ISRI

A:Cross-references: SGD:S0006310; MIPS:YPR106w

A:Map position: 16R

Query Match 9.2%; Score 70.5; DB 2; Length 443;

Best Local Similarity 22.7%; Pred. No. 14;

Matches 22; Conservative 18; Mismatches 32; Indels 25; Gaps 4;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:39:39 ; Search time 11 Seconds
(without alignments)

506.874 Million cell updates/sec

Title: US-09-673-400a-38

Perfect score: 763
Sequence: 1 ARAPFLDMRFRRLSADPHA.....YRGFLAEKRVFGWEDEDF 144

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	335	43.9	146	1 LD01_HUMAN	Q05751 homo sapien
2	74	9.7	414	1 YW42_MYCTU	Q10533 mycobacteri
3	74	9.7	970	1 NAC1_MOUSE	P70424 mus musculu
4	69	9.0	348	1 RMO3_HUMAN	P09001 homo sapien
5	69	9.0	1487	1 ICP4_HSVK	P17473 equine herp
6	68.5	9.0	1515	1 YCF1_YEAST	P39109 saccharomyc
7	68	8.9	514	1 OAS1_HUMAN	Q15646 homo sapien
8	67.5	8.8	316	1 CYR1_STRCO	Q02055 streptomyce
9	67.5	8.8	971	1 NAC1_RAT	Q01728 rattus norv
10	66.5	8.7	430	1 FTST_MYCLE	Q33010 mycobacteri
11	66.5	8.7	914	1 PERT_MOUSE	P35419 mus musculu
12	66.5	8.7	914	1 PERT_RAT	P14650 rattus norv
13	66	8.7	488	1 SYK_MYCFE	Q49158 mycoplasma
14	66	8.7	875	1 YC78_MYCTU	Q11042 mycobacteri
15	66	8.7	989	1 PPOC_LEUME	P94892 leucostoc
16	66	8.7	159	1 SECT_YEAST	P11075 saccharomyc
17	65.5	8.6	159	1 LARX_PAPSO	P19825 papaver som
18	65.5	8.6	822	1 HIFA_MOUSE	Q09121 homo sapien
19	65.5	8.6	4829	1 BIR6_HUMAN	P04223 mus musculu
20	65	8.5	369	1 HAIK_MOUSE	Q09K50 chlamydia m
21	65	8.5	819	1 LON_CHLMU	P37893 caulobacter
22	65	8.5	863	1 AMPN_CAVCR	P13607 drosophila
23	65	8.5	1041	1 ATNA_MOUSE	P28925 equine herp
24	65	8.5	1487	1 ICP4_HSVK	P22414 candida tro
25	64.5	8.5	906	1 FOX2_CANTR	Q14089 schizosach
26	64.5	8.5	910	1 TMR2_SCHPO	P37276 drosophila
27	64.5	8.5	4639	1 DYRC_MOUSE	P35922 mus musculu
28	64	8.4	614	1 FMRI_MOUSE	O84348 chlamydia t
29	64	8.4	819	1 LON_CHLMU	P53552 saccharomyc
30	64	8.4	1597	1 RLRL_YEAST	P20809 homo sapien
31	63.5	8.3	139	1 IL11_HUMAN	O09689 schizosach
32	63.5	8.3	395	1 IF5_SCHPO	P30724 graciilaria
33	63.5	8.3	416	1 G3PA_GRAVE	

34	63.5	8.3	630	1 TNPB_STAAT	P06697 staphylococ
35	63.5	8.3	874	1 UIS2_EBV	P03193 Epstein-Bar
36	63.5	8.3	1698	1 Y076_HUMAN	Q14999 homo sapien
37	63	8.3	619	1 Y016_CHLPN	O926P3 chlamydia p
38	63	8.3	724	1 MALO_MYCTU	O53932 mycobacteri
39	63	8.3	819	1 LON_CHLPN	O928F4 chlamydia p
40	63	8.3	973	1 NAC1_HUMAN	P32418 homo sapien
41	63	8.3	1005	1 YCF1_OENBE	P31563 oenocthera b
42	62.5	8.2	472	1 SACH_BACAM	P21130 bacillus am
43	62.5	8.2	671	1 RIP_HUMAN	Q13546 homo sapien
44	62	8.1	303	1 YE78_METJA	O58873 methanococc
45	62	8.1	348	1 RMO3_MOUSE	O99895 mus musculu

ALIGNMENTS

RESULT 1	ID	LD01_HUMAN	STANDARD:	PRT:	146 AA.
AC	Q05751:				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	LD0C1 protein (Leucine zipper protein down-regulated in cancer cells).				
GN	LD0C1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=99330357; PubMed=10403563;				
RA	Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T.,				
RA	Yamaguchi K.;				
RT	Identification of a novel gene, LD0C1, down-regulated in cancer cell				
RT	lines.;				
RL	Cancer Lett. 140:227-234(1999).				
CC	- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN THE DEVELOPMENT AND/OR				
CC	PROGRESSION OF SOME CANCERS.				
CC	- SUBCELLULAR LOCATION: Nuclear.				
CC	- TISSUE SPECIFICITY: OBITUOUSLY EXPRESSED WITH HIGH LEVELS IN				
CC	BRAIN ANT THYROID AND LOW EXPRESSION IN PLACENTA, LIVER AND				
CC	LEUCOCYTES. EXPRESSED AS WELL IN SIX OF THE SEVEN HUMAN BREAST				
CC	CANCER CELL LINES EXAMINED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; AB019527; BAA34364.1; -				
KW	Nuclear protein.				
FT	DOMAIN 132 143				
FT	SEQUENCE 146 AA: 16968 MW: 3D02813B2DE52DBE CRC64:				

Query Match 43.96; Score 335; DB 1; Length 146;
Best Local Similarity 59.64; Pred. No. 6.5e-30;
Matches 62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;

QY	40	KALLAGLRPARMRNDPPEPEFDGDTDLPEFIYOTCSYMPVDENTSNDALKYFL	99
DB	37	RASLRLQVYRPPS---CPVPEPEFNGESSLPPEFYQTASTYLVNRRFCNDKVAFL	92
QY	100	ITRLTGPALQVPIYIRKESPLNDYRGFLAEKRVFGWEDEDF	143
DB	93	ISLTGRAEVYVYIENDSYILDYRAFLDEMKQCRGWDDDD	136

```

RESULT 2
YM42_MYCITU          STANDARD;      PRT;       414 AA.
ID   YM42_MYCITU
AC   O10523;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DE   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 44.6 kDa protein RV2242.
GN   RV2242 OR MT2302 OR MYCIV427.23.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RX   NCBI_TaxID=1773;
RN   [1]
RP   STRAIN FROM N.A..
RC   SEQUENCE-H3TRV;
RK   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eigmler K., Gass S., Barry C.E. III, Tekala F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT   Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence."
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A..
RC   STRAIN=CDC 151 / Oshkosh;
RA   Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains".
RL   Submitted (APR-2001) to the EMBL/Genbank/DDBJ databases.
CC   -I- SIMILARITY: BELONGS TO THE CAR FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outpost.
CC   The European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: Z70692; CA94663.1; -.
DR   EMBL: AE007074; AAA46386.1; -.
DR   TIGR: MT2302; -.
DR   Tuberculist; RV2242; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 414 AA; 44637 MW; F454D4339771F73 CRC64;
-----
Query Match           9.7%; Score 74; DB 1; Length 414;
Best Local Similarity 25.8%; Pred. NO. 1.5;
Matches 24; Conservative 17; Mismatches 32; Indels 20; Gaps 4;
OY    8 MFRRRLASADPHATQNSAARGTGMDGRVO--LMKALLAGPLRP-----AARRW----- 54
DB    141 LKYSROLATPAAAYADAADAEARGCTMDSNRBAASVDVAVNGDTGPELLSAALAMNDTPAP 200
OY    55 -----RNPIFP--ETFGDDTDRLRLEPFIVQCFS 80
DB    201 ATVLVGTPAPGPNGSNSDSDSERASODVDPTAA 233

RESULT 3
ID   NACL_MOUSE
AC   P70414;
STANDARD;      PRT;       970 AA.

```

DT	01-NOV-1997	(Rel. 35, Created)
DT	01-NOV-1997	(Rel. 35, Last sequence update)
DT	01-MAR-2002	(Rel. 41, Last annotation update)
DE	Sodium/calcium exchanger 1 precursor (Na(+)/Ca(2+)-exchange protein 1).	
GN	SLC8A1 OR NCX.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6;	
RX	MEDLINE=96250070; PubMed=8659820;	
RA	Kim I., Lee C.O.;	
RT	*Cloning of the mouse cardiac Na(+)-Ca2+ exchanger and functional expression in Xenopus oocytes.*;	
RL	Ann. N.Y. Acad. Sci. 779:126-128(1996).	
CC	- FUNCTION: RAPIDLY TRANSPORTS CA2+ DURING EXCITATION-CONTRACTION COUPLING. CA(2+) IS EXTRUDED FROM THE CELL DURING RELAXATION SO AS TO PREVENT OVERLOADING OF INTRACELLULAR STORES.	
CC	- ENZYME REGULATION: BY ATP..	
CC	- INTRACELLULAR LOCATION: Integral membrane protein. Plasma membrane.	
CC	- TISSUE SPECIFICITY: CARDIAC SARCOLEMA.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@sdb-sib.ch).	
CC	-----	
DR	EMBL; U70033; AAB46708.1; -.	
DR	MGD; MGI:107956; Slc8a1.	
DR	InterPro; IPR003644; Calx_beta.	
DR	InterPro; IPR002613; Na_Ca_Ex.	
DR	Pfam; PF01699; Na_Ca_Ex.2.	
DR	PRINTS; PR01259; NACACXCHNGR.	
KW	SMART; SMO0237; Calx_beta; 2.	
KW	Transport; Symport; Transmembrane; Glycoprotein; Phosphorylation; Signal; Calmodulin-binding; Calcium transport; Repeat.	
FT	CHAIN	1 32
FT	DOMAIN	33 970
FT	TRANSMEM	72 71
FT	DOMAIN	94 93
FT	TRANSMEM	134 133
FT	DOMAIN	156 155
FT	TRANSMEM	168 188
FT	DOMAIN	189 199
FT	TRANSMEM	200 222
FT	DOMAIN	223 225
FT	TRANSMEM	226 249
FT	DOMAIN	250 769
FT	TRANSMEM	770 789
FT	DOMAIN	790 796
FT	TRANSMEM	797 819
FT	DOMAIN	820 821
FT	TRANSMEM	822 840
FT	DOMAIN	841 871
FT	TRANSMEM	872 892
FT	DOMAIN	893 903
FT	TRANSMEM	904 924
FT	DOMAIN	925 941
FT	TRANSMEM	942 958
FT	DOMAIN	959 970
FT	TRANSMEM	251 270
FT	DOMAIN	138 178
FT	REPEAT	1407 478
FT	REPEAT	539 609
FT	REPEAT	839 875
FT	DOMAIN	236 239
FT	POLY-PHE.	

Query Match 9.7%; Score 74; DB 1; Length 970;
Best Local Similarity 27.0%; Pred. No. 4.4;
Matches 34; Conservative 20; Mismatches 40; Indels 32; Gaps 7;

FT DOMAIN 689 692 POLY-GLT.
FT DOMAIN 756 760 POLY-ASP.
FT MOD_RES 389 389 PHOSPHORYLATION (POTENTIAL).
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 970 AA: 108035 MW: 65500 DB07F2B6602 CRC64;

Query Match 9.7%; Score 74; DB 1; Length 970;
Best Local Similarity 27.0%; Pred. No. 4.4;
Matches 34; Conservative 20; Mismatches 40; Indels 32; Gaps 7;

DB 547 SESIGIMEYKV-----LRTSGARGNVILIPY-KTIEGTARGGCEDEFDCCGEPEFO 595
OY 26 AEARCTMDGRVQMLKALLAGPLRPARARWRNPPEPEFDG-----DTRLPPE- 74
DB 75 ---IVQTSYMEVDENTFSNDALKVTLRLTGLPALQWVIRYIKRESPLNDYREGFLAE 131
DB 596 NDEIVKTTISVKVYIDDEYEKRN--TFEL-ELGEPRL--VENSEKALLNLDELGGFTLT 648
OY 132 MKRVFG 137
DB 649 GKEMTG 654

RESULT 4
RM03_HUMAN
ID RM03_HUMAN STANDARD: PRT: 348 AA.
AC P09001.
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Mitochondrial 60S ribosomal protein L3.
GN MRPL3 OR MR13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP -SEQUENCE FROM N.A.
RA MEDLINE=88067705; PubMed=2891103;
RA Ou J.-H., Yen T.S.B., Wang Y.-F., Kam W.K., Rutter W.;
RT "Cloning and characterization of a human ribosomal protein gene with
RT enhanced expression in fetal and neoplastic cells.";
RL Nucleic Acids Res. 15:8919-8934(1987).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X06323; CAA29639.1; -
CC PIR: S01607; R5HUL3.
DR InterPro: IPR000597; Ribosomal_L3.
DR Pfam: PF00297; Ribosomal_L3; 1.
DR ProDom: PD001374; Ribosomal_L3; 1.
DR PROSITE: PS00474; RIBOSOMAL_L3; 1.
KW Ribosomal protein; Mitochondrion.
SO SEQUENCE 348 AA: 38632 MW: 60655 DB06DB72E32 CRC64;

Query Match 9.0%; Score 69; DB 1; Length 348;
Best Local Similarity 29.1%; Pred. No. 4.4;
Matches 23; Conservative 10; Mismatches 36; Indels 10; Gaps 3;

OY 21 TQNSAARCTMDGRVQMLKALLAGPLRPARARWRNPPEPEFDG-DGDTRLPPEFTVOTC 79
DB 278 TKHNIIVYNGSVGHNKCLVKKDSKL-PAYKDLGKNLPPEYTFPDGDEELPDDL----- 332

OY 80 SYMEVDENTFSNDALKVTF 98
DB 333 ----YDENWQPGAPSITE 347

RESULT 5
ICP4_HSVK
ID ICP4_HSVK STANDARD: PRT: 1487 AA.
AC P17473.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Trans-acting transcriptional protein ICP4 (155 kDa immediate-early
DE protein).
GN IE OR 64.
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_Taxid=10329;
RN [1]
RP -SEQUENCE FROM N.A.
RA MEDLINE=89370304; PubMed=2549711;
RA Grundy F.J., Baumann R.P., O'Callaghan D.J.;
RT "DNA sequence and comparative analyses of the equine herpesvirus type
RT 1 immediate early gene.";
RT Virology 172:223-236(1989).
RN [2]
RP -SEQUENCE OF 1432-1487 FROM N.A.
RX MEDLINE=90064773; PubMed=2555546;
RX Grundy F.J., Baumann R.P., O'Callaghan D.J.;
RA "Mapping the termini and intron of the spliced immediate-early
RT transcript of equine herpesvirus 1.";
RL J. Virol. 63:5101-5110(1989).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND ADAPTING TO INFECTED CELLS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J04366; AAA45089.1; -
CC EMBL: M30498; AAA66554.1; -
CC PIR: A33764; EDBEEL.
DR HSPSP: P04002; IWEA.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 181 213 SER-RICH.
FT DOMAIN 922 931 ARG/LYS-RICH (BASIC).
SO SEQUENCE 1487 AA: 154716 MW: 044E39A570608A6B CRC64;

Query Match 9.0%; Score 69; DB 1; Length 1487;
Best Local Similarity 24.6%; Pred. No. 26;
Matches 33; Conservative 15; Mismatches 48; Indels 38; Gaps 5;

OY 10 FRRRLSADPHATQRSAAARG--TWDGAVQMLKALLAGPLRPARARWRNPPEPEFDGD 67
DB 1035 WRPALTEPPALATIAACSGPRARD-ARGELAAAGPLARRA-AMHQLPDPED----- 1088
OY 68 TDRLPPEFTVOTCSYMEVDENTFSNDALKVTLRLTGLPALQWVIRYIKRESPLN 123
DB 1089 -----YKVVVLYSPLQEDDLGLGAPASPGSGSREPLMS 1122
OY 124 DYRGFLAMKRVFG 137

DB 1123 DLKGLSALLAALG 1136

RESULT 6

YCFL_YEAST STANDARD: PRT: 1515 AA.

ID YCFL_YEAST 003905;

AC P39109; 003905;

DT 01-FEB-1995 (Rel. 31, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Metal resistance protein YCFL (Yeast cadmium factor 1).

GN YCFL OR YDR135C OR YD9302.11C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H9;

RX MEDLINE=94357936; PubMed=7521334;

RA Sczyzka M.S., Memme J.A., Moye-Rowley S.W., Thiele D.J.;

RT "A yeast metal resistance protein similar to human cystic fibrosis transmembrane conductance regulator (CFTR) and multidrug resistance-associated protein."

RT J. Biol. Chem. 269:22853-22857(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=SZ88C / AB972;

RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;

RT Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.

RN [3]

RP FUNCTION.

RX MEDLINE=20253522; PubMed=10790694;

RA Petrovic S., Pascolo L., Gallo R., Cupelli F., Ostrow J.D., Goffeau A., Tiribelli C., Bruschi C.V.;

RT "The products of YCFL and YL1015w (BPT1) cooperate for the ATP-dependent vacuolar transport of unconjugated bilirubin in Saccharomyces cerevisiae."

RT Yeast 16:561-571(2000).

CC -1- FUNCTION: COOPERATES FOR THE ATP-DEPENDENT VACUOLAR TRANSPORT OF BILIRUBIN.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL: L35237; AAA50353.1;

DR EMBL: Z48179; CA88217.1;

DR HSP: P13569; INED.

DR SGD: S0002542; YCFL.

DR InterPro: IPR003593; AAA.

DR InterPro: IPR001140; ABC_transporter_tmam.

DR InterPro: IPR003439; ABC_transporter.

DR Pfam: PF00664; ABC_membrane_2.

DR Pfam: PF00005; ABC_tran_2.

DR SMART: SM00382; AAA_1.

DR PROSITE: PS00211; ABC_TRANSPORTER_2.

DR ATP-binding; Transmembrane; Glycoprotein; Transport;

KW Cadmium resistance; Phosphorylation.

KW DOMAIN 1

FT TRANSMEM 33 53 1 (BY SIMILARITY).

FT DOMAIN 54 73 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 74 94 2 (BY SIMILARITY).

FT DOMAIN 95 99 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 100 120 3 (BY SIMILARITY).

FT DOMAIN 121 130 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 131 151 4 (BY SIMILARITY).

FT DOMAIN 152 169 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 170 190 5 (BY SIMILARITY).

FT DOMAIN 191 278 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 279 299 6 (BY SIMILARITY).

FT DOMAIN 300 345 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 346 366 7 (BY SIMILARITY).

FT DOMAIN 367 422 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 423 443 8 (BY SIMILARITY).

FT DOMAIN 444 446 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 447 467 9 (BY SIMILARITY).

FT DOMAIN 468 530 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 531 551 10 (BY SIMILARITY).

FT DOMAIN 552 572 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 573 593 11 (BY SIMILARITY).

FT DOMAIN 594 943 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 944 964 12 (BY SIMILARITY).

FT DOMAIN 965 1001 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1002 1023 13 (BY SIMILARITY).

FT DOMAIN 1024 1066 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 1067 1087 14 (BY SIMILARITY).

FT DOMAIN 1088 1088 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1089 1109 15 (BY SIMILARITY).

FT DOMAIN 1110 1180 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 1181 1201 16 (BY SIMILARITY).

FT DOMAIN 1202 1205 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1206 1226 17 (BY SIMILARITY).

FT DOMAIN 1227 1515 CYTOPLASMIC (BY SIMILARITY).

FT NP_BIND 663 670 ATP (POTENTIAL).

FT NP_BIND 856 863 ATP (POTENTIAL).

FT NP_BIND 1306 1313 ATP (POTENTIAL).

FT MOD_RES 908 908 PHOSPHORYLATION (BY CAPK) (POTENTIAL).

FT MUTAGEN 713 713 MISSING: LOSS OF FUNCTION.

FT MUTAGEN 908 908 S->A: LOSS OF FUNCTION.

FT CONFLICT 680 680 L->R (IN REF. 1).

SO SEQUENCE 1515 AA; 171120 MW; 30F92FEDBFAF60431 CRC64;

Query Match

Best Local Similarity 23.4%; Score 68.5; DB 1; Length 1515;

Matches 29; Conservative 19; Mismatches 35; Indels 41; Gaps 5;

OY 55 RNP1PPEP-----FDGDTDRLEPIVOTCSYFVDENFNSDALKVFLITPLNG 105

DB 1035 RAMPTEFTPIGRILNRSDNYKVDALGRFSQFV-----NAVVTITIVICA 1087

OY 106 PALOW---VIP-----YIR-----KESPLNDYRCFLAEKRRVSGWEE 140

DB 1088 TWTQGFIIILPSVFIYVOQVYLRSRELRLDSTIRSPYISHFOTGLGATVNGYSQ 1147

OY 141 DEDE 144

DB 1148 QKRPF 1151

RESULT 7

OASL_HUMAN STANDARD: PRT: 514 AA.

ID OASL_HUMAN 015646; 075686; Q9Y6K7; Q9Y6K6;

AC Q15646; 075686; Q9Y6K7; Q9Y6K6;

DT 01-NOV-1997 (Rel. 35, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 59 kDa 2'-5'-oligoadenylate synthetase like protein (p59 OASL)

DE (p59OASL) (Thyroid receptor interacting protein 14) (TRIP14).

GN OASL OR TRIP14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM P56).

RP MEDLINE=98391734; PubMed=9722630;

RA Hartmann R., Olsen H.S., Wilder S., Joergensen R., Justesen J.;

RT "p59OASL, a 2'-5' oligoadenylate synthetase like protein: a novel human gene related to the 2'-5' oligoadenylate synthetase family."

RL Nucleic Acids Res. 26:4121-4127(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS P56 AND P30).
 RC TISSUE-Monocytes;
 RX MEDLINE-99041549; PubMed-9826176;
 RA Reboullet D., Marie I., Hovanessian A.G.;
 RT "Molecular cloning and characterization of two related and interferon-
 induced 56-kDa and 30-kDa proteins highly similar to 2'-5'-
 oligoadenylate synthetase."
 RL Eur. J. Biochem. 257:319-330(1998).
 RP [3]
 RA SEQUENCE FROM N.A.
 RL White S.;
 RP Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 260-416 FROM N.A. (ISOFORM P56).
 RX MEDLINE-95295737; PubMed-7776974;
 RA Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;
 RT "Two classes of proteins dependent on either the presence or absence
 of thyroid hormone for interaction with the thyroid hormone
 receptor."
 RL Mol. Endocrinol. 9:243-254(1995).
 CC -1- FUNCTION: DOES NOT HAVE 2'-5'-OAS ACTIVITY, BUT BINDS DOUBLE-
 STRANDED RNA AND DNA.
 CC -1- SUBUNIT: SPECIFICALLY INTERACTS WITH THE LIGAND BINDING DOMAIN OF
 THE THYROID RECEPTOR (TR). TRIP14 DOES NOT REQUIRE THE PRESENCE OF
 THYROID HORMONE FOR ITS INTERACTION.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: P56 (SHOWN HERE) AND P30; MAY BE
 PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH THE HIGHEST
 LEVELS IN PRIMARY BLOOD LEUKOCYTES AND OTHER HEMATOPOIETIC SYSTEM
 TISSUES, COLON, STOMACH AND TO SOME EXTENT IN TESTIS.
 CC -1- INDUCTION: BY INTERFERONS.
 CC -1- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 UNIQUTIN-LIKE DOMAINS.
 CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 386
 TO 416 DUE TO A FRAMESHIFT.
 CC -1- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF MOUSE OASL.
 CC -----
 CC -This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ225089; CAI12396.1; -;
 DR EMBL: AF063611; AAD28541.1; -;
 DR EMBL: AF063612; AAD28542.1; -;
 DR EMBL: Z93097; -; NOT_ANNOTATED_CDS.
 DR EMBL: LA0387; AAC41733.1; ALT_FRAME.
 DR HSSP: P02248; 10BI.
 DR MIM: 603281; -;
 DR InterPro: IPR001797; 25A_synth.
 DR InterPro: IPR001201; PAP_25A_core.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; ubiquitin, 2.
 DR SMART: SM00213; UBO; 2.
 DR PROSITE: PS00832; 25A_SYNTH_1; 1.
 DR PROSITE: PS00833; 25A_SYNTH_2; 1.
 DR PROSITE: PS0152; 25A_SYNTH_3; 1.
 DR PROSITE: PS0053; UNIQUTIN_2; 1.
 KM RNA-binding; Interferon induction; Repeat.
 FT DOMAIN 354 433
 FT UNIQUTIN-LIKE 1.
 FT DOMAIN 434 509
 FT UNIQUTIN-LIKE 2.
 FT VARSPPLIC 220 255
 FT YVARSPPRALPLVLALELITIAMENGTEEDENFM -> A
 FT HHGSGRPHORGRVOMGRCSEGLFVPTGLLL (IN
 FT ISOFORM P30).
 FT VARSPPLIC 256 514
 FT MISSING (IN ISOFORM P30).
 FT CONFLICT 26 38
 FT HREKKEVLDVAVR -> TVEGRGARCA (IN REF.
 FT 2).
 FT CONFLICT 89 89
 FT S -> T (IN REF. 2; AAD28542).

FT CONFLICT 95 113
 FT OPAKHKDVLRIKTMW -> PGGSQASQRCSEADMENH
 FT (IN REF. 2).
 FT A -> S (IN REF. 2).
 FT CONFLICT 223 223
 FT Y -> I (IN REF. 2).
 FT CONFLICT 317 317
 FT I -> T (IN REF. 2).
 FT CONFLICT 321 321
 FT O -> L (IN REF. 2).
 FT CONFLICT 324 324
 FT NP -> KG (IN REF. 4).
 FT CONFLICT 341 342
 FT S -> T (IN REF. 2; AAD28541).
 FT CONFLICT 445 445
 FT SEQUENCE 514 AA; 59226 MW; 4DBBB655D9EA003E CRC64;
 Query Match 8.9%; Score 68; DB 1; Length 514;
 Best Local Similarity 24.3%; Pred. No. 9.2;
 Matches 37; Conservative 24; Mismatches 55; Indels 36; Gaps 10;
 QY 7 DMFRRRLSADPHATONSASARCTMDG-RVQIMKALLA-GILRPAARMNPPIPETE 64
 DB 125 DLRMGEV--PDALV-FTIOTGTAPITVIVPAVALGSPSP-----NSQPPPEVY 174
 QY 65 DGOTDLRPEIVOTC-----SYMFVDTEFSDALKYFTLRLTGRPALOWIPIYR 116
 DB 175 VS-----LIRACGGPGNFCPSFSELQRFVYKHRTKLSLL-RL--VKHWYQOIVK 222
 QY 117 KESPLNDYRGFLAEMKRVFGW---EDEDEF 144
 DB 223 ARSPPRALPLVLALELITIAMENGTEEDENF 254
 RESULT 8
 CYRK_STRCO STANDARD; PRT; 316 AA.
 ID CYRK_STRCO 002055;
 AC 002055;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Actinorhodin polyketide synthase bifunctional cyclase/dehydratase
 DE (actI ORF4) (ACTVII).
 GN SCBAC2861.16.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID:1902;
 RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE-92406871; PubMed-1527048;
 RA Fernandez-Moreno M.A., Martinez E., Boto L., Hopwood D.A.,
 Malpartida F.;
 RT "Nucleotide sequence and deduced functions of a set of cotranscribed
 RT genes of Streptomyces coelicolor A3(2) including the polyketide
 RT synthase for the antibiotic actinorhodin."
 RL J. Biol. Chem. 267:19278-19290(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Warren T., Harris D., Cerdano A.M., Parkhill J., Barrell B.G.,
 Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IS NEEDED FOR CORRECT CYCLIZATION OF THE OLIGOPEPTIDE
 CC LEADING TO ISOCHROMANONE QUINONE FORMATION.
 CC -1- PATHWAY: BIOSYNTHESIS OF POLYKETIDE ANTIBIOTIC ACTINORHODIN.
 CC -----
 CC -This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63449; CAA45046.1; -;
 DR EMBL: AL593842; CAC44203.1; -;
 DR PIR: S25843; S25843.
 KM Antibiotic biosynthesis.

FT CONFLICT 402 402 P -> A (IN REF. 1).
SQ SEQUENCE 971 AA; 108184 MW; EC456CE3AC6A69 CRC64;
Query Match 8.8%; Score 67.5; DB 1; Length 971;
Best Local Similarity 27.6%; Pred. No. 23;
Matches 35; Conservative 21; Mismatches 38; Indels 33; Gaps 8;

26 AEARGTGDRVQMLKALLAGPLRPARRRNPDPPEPTFG-----DTDRLPPE- 74
DB SESIGIMEVK-----LRTSGANGVITIP-KTTEGTARGCGDEPDTGCELEFQ 595
OY 75 ---IVQTSYMEVDENTSNDALKTFLITRLTGALQWVPIRKESPLNDYGF-IA 130
DB 596 NDEIVKTIISVYIDEEYEKNK---TFPI-ETGEPRL---VEMSEKALLNELGGFTLT 648
OY 131 EMKRVFG 137
DB 649 EGKKTG 655

RESULT 10
FTSY_MYCLE
ID FTSY_MYCLE STANDARD: PRT: 430 AA.
AC 033010;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein ftsy homolog.
GN FTSY OR ML1628 OR MLCB250.02.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-TN:
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Houtby L., Hornsby T., Jagels K., Isacstrix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Stevens K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: FUNCTIONAL HOMOLOGY OF A SUBSET OF PROTEINS AT THE
CC THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
CC MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 297369; CAB10596.1; -
DR EMBL: AL583922; GAC30579.1; -
DR HSSP: P10121; 1FTS.
DR Leptoma; ML1628;
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001687; ATP-GTP_A.
DR InterPro: IPR000897; SRP54.
DR Pfam: PF00448; SRP54; 1.
DR Pfam: PF02881; SRP54.N; 1.
DR ProDom: PD000819; SRP54; 1.
DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding; Membrane;
KW Cell division; Complete proteome.
FT NP_BIND 238 245 GTP (BY SIMILARITY).
FT NP_BIND 320 324 GTP (BY SIMILARITY).
FT NP_BIND 382 385 GTP (BY SIMILARITY).
SQ SEQUENCE 430 AA; 45354 MW; E9F78589AAB2C598 CRC64;
Query Match 8.7%; Score 66.5; DB 1; Length 430;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 40; Conservative 19; Mismatches 57; Indels 47; Gaps 6;

8 MRRERR-----LSADPHATQNSAEKRGTMGRVQMLKALLAGPLRPARRRNPPI 59
DB 26 VYRRRRRLSLPSSLSGSGASDRS---GSYTAASGTFESQALPVQPADRI----- 74
OY 60 FPEFDGDPRLPEFTVQCSYMFVDENTSNDALKTFLITRLTGALQWVPIRKES 119
DB 75 -----DTGELPANGDQATVPRDSRPHISDVLPESELITSPDEP--EAAVPHIDATV 125
OY 120 P---LLNDYRGFLAEKRVFG-----WEDED 143
DB 126 PEGRLDLRLRGRLARTQNAFGRSILGLVGDLEDSDWQELBD 168

RESULT 11
PERT_MOUSE
ID PERT_MOUSE STANDARD: PRT: 914 AA.
AC P35419;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thyroid peroxidase precursor (P35419.1.8) (TPO).
GN TPO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thyroid;
RX MEDLINE=93154601; PubMed=7916704;
RA Kotani T., Umeki K., Yamamoto I., Takeuchi M., Takeuchi S.,
RA Nakayama T., Ohtaki S.;
RT "Nucleotide sequence of the cDNA encoding mouse thyroid peroxidase.";
RL Gene 123:289-290(1993).
CC -1- FUNCTION: IODINATION AND COUPLING OF THE HORMONAL TYROSINES
CC IN THYROGLOBULIN TO YIELD THE THYROID HORMONES T(3) AND T(4).
CC -1- CATALYTIC ACTIVITY: Iodide + H(2)O(2) = Iodine + 2 H(2)O.
CC -1- COFACTOR: HEME (PROTOPORPHYRIN IX).
CC -1- PATHWAY: THYROID HORMONE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Membrane.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60703; CAA43114.1; -
DR PIR: JN0550; JN0550.
DR HSSP: P05164; 1CXP.
DR MGD: MGT:98813; Tpo.
DR InterPro: IPR002007; Anim_peroxidase.
DR InterPro: IPR000157; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001536; Peroxidase_3.
DR InterPro: IPR002016; Peroxidase.

DR Interpro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF03098; An_peroxidase; 1.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF00084; sushi; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR SMART: SM00032; CCP; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00435; PEROXIDASE_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR Oxidoreductase; Peroxidase; Heme; Transmembrane; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 914
 FT ACT_SITE 233 233
 FT ACT_SITE 384 384
 FT ACT_SITE 482 482
 FT TRANSMEM 835 859
 FT CARBOHYD 123 123
 FT CARBOHYD 271 271
 FT CARBOHYD 299 299
 FT CARBOHYD 334 334
 FT CARBOHYD 603 603
 FT SEQUENCE 914 AA; 101342 MW; 59569A0B71F3DD01 CRC64;

Query Match 8.7%; Score 66.5; DB 1; Length 914;
 Best Local Similarity 30.9%; Pred. No. 27;
 Matches 21; Conservative 12; Mismatches 32; Indels 3; Gaps 2;

Oy 61 PETFGDTRLPETIVOCYMFVDFNTFSNDALKVFL-ITRLTGPALQWIPYIRKE 118
 Db 80 PESTSGAISRAELMETSIQWKREQSFTDALSADILATIANLSG-CLPFLMPPRCDD 138
 Oy 119 SPLNDYR 126
 Db 139 TCLANKYR 146

RESULT 12

PERT_RAT STANDARD; PRT; 914 AA.
 AC P14650;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thyroid peroxidase precursor (EC 1.11.1.8) (TPO).
 GN TPO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90045972; PubMed-2813071;
 RA Derwahl M., Seto P., Rapoport B.;
 RT "Complete nucleotide sequence of the cDNA for thyroid peroxidase in
 FRTL-5 rat thyroid cells";
 RL Nucleic Acids Res. 17:8380-8380(1989).
 RN [2]
 RP SEQUENCE OF 145-914 FROM N.A.
 RX MEDLINE-90114171; PubMed-2691880;
 RA Isozaki O., Kohn L.D., Kozak C.A., Kimura S.;
 RT "Thyroid peroxidase: rat cDNA sequence, chromosomal localization in
 mouse, and regulation of gene expression by comparison to
 thyroglobulin in rat FRTL-5 cells";
 RL Mol. Endocrinol. 3:1681-1692(1989).
 CC -1- FUNCTION: IODINATION AND COUPLING OF THE HORMONOGENIC TYROSINES
 CC IN THYROGLOBULIN TO YIELD THE THYROID HORMONES T(3) AND T(4).
 CC -1- CATALYTIC ACTIVITY: Iodide + H(2)O(2) -> iodine + 2 H(2)O.
 CC -1- COFACTOR: HEME (PROTOPORPHYRIN IX).
 CC -1- PATHWAY: THYROID HORMONE BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: Membrane.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: X17396; CA35257.1; -;
 CC EMBL: M31655; AAA42265.1; -;
 CC PIR: S07047; S07047.
 CC HSSP: P05164; ICXP.
 DR Interpro: IPR002007; Anim_peroxidase.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF_CA.
 DR Interpro: IPR001536; Peroxidase_3.
 DR Interpro: IPR002016; Peroxidase.
 DR Interpro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF03098; An_peroxidase; 1.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF00084; sushi; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00032; CCP; 1.
 DR PROSITE: PS00435; PEROXIDASE_2; FALSE_NEG.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00435; PEROXIDASE_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 KW Oxidoreductase; Peroxidase; Heme; Transmembrane; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 914
 FT ACT_SITE 233 233
 FT ACT_SITE 384 384
 FT ACT_SITE 482 482
 FT TRANSMEM 835 859
 FT CARBOHYD 123 123
 FT CARBOHYD 271 271
 FT CARBOHYD 299 299
 FT CARBOHYD 334 334
 FT CARBOHYD 603 603
 FT CARBOHYD 194 195
 FT CONFLICT 198 198
 FT CONFLICT 228 228
 FT CONFLICT 592 594
 FT SEQUENCE 914 AA; 101460 MW; B700B89439E85191 CRC64;

Query Match 8.7%; Score 66.5; DB 1; Length 914;
 Best Local Similarity 30.9%; Pred. No. 27;

Matches 21; Conservative 12; Mismatches 32; Indels 3; Gaps 2;

Oy 61 PETFGDTRLPETIVOCYMFVDFNTFSNDALKVFL-ITRLTGPALQWIPYIRKE 118
 Db 80 PESTSGAISRAELMETSIQWKREQSFTDALSADILATIANLSG-CLPFLMPPRCDD 138
 Oy 119 SPLNDYR 126
 Db 139 TCLANKYR 146

RESULT 13

SVK_MYCFE STANDARD; PRT; 488 AA.
 AC O49158;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LYSRS).
 GS LYXS.
 OS Mycoplasma fermentans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 NCBI_TaxID=2115;
 RX MEDLINE=96201595; PubMed=8613394;
 RA STRAIN=PC18;
 RT Theiss P., Karpas A., Wise K.;
 RT "Antigenic topology of the P23 surface lipoprotein of Mycoplasma
 RT fermentans: differential display of epitopes results in
 RT high-frequency phase variation."
 RT Infect. Immun. 64:1800-1809(1996).
 CC -1 CATALYTIC ACTIVITY: ATP + L-Lysine + tRNA(Lys) = AMP + diphosphate
 CC + L-Lysyl-tRNA(Lys).
 CC -1 SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1 SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U50826; AAC43988.1; -
 DR HSSP: P14825; ILYL.
 DR Interpro: IPR002106; AA-trna_ligase_II.
 DR Interpro: IPR002309; tRNA-synt_2.
 DR Interpro: IPR002313; tRNA-synt_lys_2.
 DR Pfam: PF00152; tRNA-synt_2_1.
 DR Pfam: PF01336; tRNA-anti_1.
 DR PRINTS: PRO0982; TRNASYNTHLYS.
 DR PROSITE: PS00179; AA-trna_ligase_II_1; 1.
 DR PROSITE: PS00339; AA-trna_ligase_II_2; 1.
 KM "Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
 SQ SEQUENCE 488 AA; 56703 MW; 6F42D2B51413202 CRC64;
 Query Match 8.7%; Score 66; DB 1; Length 488;
 Best Local Similarity 25.7%; Pred. No. 14;
 Matches 19; Conservative 17; Mismatches 20; Indels 18; Gaps 5;
 QY 59 PEPPEFDGDTLRPEFIYQTCSYMFVDFNTFSNDALKVTELTITRLTGPALQWV-----111
 DB 138 PLPDEFHGLTDEERY-----RHRTLD--LITNPESRNTFIMKTKI-----VOMIRDYENK 186
 QY 112 IPIYKESPLANDY 125
 DB 187 LDYLEAFEPFLADY 200
 RESULT 14
 YC78_MYCTU STANDARD: PRT; 875 AA.
 AC 011042;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Hypothetical 93.4 kDa protein Rv1278.
 GN Rv1278 OR MTJ315 OR MTCY50.04C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RX MEDLINE=96295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RT Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishop W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains." to the EMBL/GenBank/DBJ databases.
 RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z77137; CAB00904.1; -
 DR EMBL: AE07006; AAK45576.1; -
 DR TIGR: MT1315; -
 DR TubercuList; Rv1278; -
 SQ Hypothetical protein; Complete proteome.
 SQ SEQUENCE 875 AA; 93350 MW; 366C580A3AAC97A2 CRC64;
 Query Match 8.7%; Score 66; DB 1; Length 875;
 Best Local Similarity 24.3%; Pred. No. 29;
 Matches 36; Conservative 21; Mismatches 61; Indels 30; Gaps 6;
 QY 12 RLSADPHATQNSAARGTMDGR-----VOLMKALIA-----GPLRPARRWRN 56
 DB 701 RKGKDLAETREHSHARVGRARRARLRSLVARRRDTRLRYVERAEHLRLGR 760
 QY 57 PIPPEPFDGDTLRPEFIYQTCSYMFVDFNTFSNDALKVTELTITRLTGPALQWVPIYR 116
 DB 761 PV-FGSPFEVEYDTLRIRSRITLDRTVPECLSGAKRGQILRIAGAL-----VA 813
 QY 117 KESP---LANDYRGF-----LAEMKRYF 136
 DB 814 KEDAVEVLIDDAIGFTDPERLAKMGVGF 841
 RESULT 15
 RPOC_LEUME STANDARD: PRT; 989 AA.
 AC P94892;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, last sequence update)
 DT 15-DEC-1998 (Rel. 37, last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 DE beta' chain) (RNA polymerase beta' subunit) (Fragment).
 GN RPOC.
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Leuconostoc.
 NCBI_TaxID=1245;
 RX MEDLINE=97016803; PubMed=8863429;

RT Morse R Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;
RA 'Analysis of the beta' subunit of DNA-dependent RNA polymerase does
RT not support the hypothesis inferred from 16S rRNA analysis that
RT Oenococcus oeni (formerly Leuconostoc oenos) is a tachytelic
RL (fast-evolving) bacterium."
RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (n).
CC -I- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X95810; CAA65077.1; -.
DR InterPro; IPR000722; RNA_POL_A.
DR Pfam; PF00623; RNA_POL_A; 1.
RW Transferase; DNA-directed RNA polymerase; Transcription.
FT NON_TER 1 1
FT ION_TER 989 989
SQ SEQUENCE 989 AA; 110185 MW; 395E91BE46F43CDC CRC64;

Query Match 8.7%; Score 66; DB 1; Length 989;
Best Local Similarity 22.7%; Pred. NO. 34;
Matches 44; Conservative 18; Mismatches 68; Indels 64; Gaps 7;

QY 1 ARAPIIDMRFRRLRSADDPATQR-----NSAEARGT 31
DB :|::|||::|||
G12 SOVAATVTKRFGCLTTDSERYORTVEITKAKDIIOQLIESFEPTNPLFMQDSGAAGN 671
QY 32 MDGRVOL--MKALLAGP-----LRPAARRWRNPDPETF-----DGDT 68
DB :|||::|::|::|::|::|::|::|::|::|:
G12 ISNFQVLQAGMGLMAGPGKIIELPVTANFRGLIVMEMFIETHGARGMGSPTALKTANS 731
QY 69 DRLPEPIVOTCSYMYVDENTFSDNAIKATFTLTTLTGALQWVIYIRKESPLNDIGCF 128
DB :|::|::|::|::|::|::|::|::|:
G73 GYLTRRLVDAADVAYIRE--FDNDSDR-----GVAAQAIMDGTSVPEELYYRIIGR 780
QY 129 LAEMKRVFGEDE 142
DB :|||::|::|::|::|::|::|::|::|:
G129 YAKSVFDEPTE 793